



UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FACULTADE DE FILOLOXÍA

DEPARTAMENTO DE LINGUA ESPAÑOLA

**ETIQUETAXE E DESAMBIGUACIÓN
AUTOMÁTICAS EN GALEGO:
O SISTEMA XIADA**

Tese de doutoramento

EVA MARÍA DOMÍNGUEZ NOYA

2013

Etiquetaxe e desambiguación automáticas en galego: o sistema XIADA

Tese de doutoramento presentada no Departamento de Lingua Española, da Facultade de Filoloxía, da Universidade de Santiago de Compostela, por Eva María Domínguez Noya e dirixida polos profesores Guillermo Rojo e M^a Sol López Martínez.

V. e Pr. do Director

V. e Pr. da Directora

Guillermo Rojo

M^a Sol López Martínez

Eva María Domínguez Noya

Para Josefa e Sebastián

A meus pais

Agradecementos

Quero expresar o meu agradecemento a todas aquelas persoas que dun xeito ou outro contribuíron a que este traballo chegase a porto, e moi especialmente a

Guillermo Rojo, por terme introducido no mundo da lingüística computacional, pola súa acertada guía e, sobre todo, por estar sempre aí;

Marisol López, pola súa constante presenza e polo seu inestimable apoio e guía;

Mario Barcala, compañeiro laboral de alegrías e tribulacións, por ter sempre unha solución para os múltiples requisitos que lle solicitaba para o sistema, mais tamén porque sen o seu traballo a maior parte dos recursos aquí descritos non poderían realizarse nin estar a disposición pública;

O Centro Ramón Piñeiro para a Investigación en Humanidades, por facilitarme o acceso aos datos necesarios para poder realizar este traballo e por confiar en min para o desenvolvemento lingüístico dos recursos aquí descritos;

Os numerosos bolseiros das salas de Lingüística e Termigal, por soportar estoicamente nestes anos que durou a elaboración da tese as miñas diatribas, en especial, Susana, Paula, Laura, Inés, Chus, Sandra, Irene...

Xesús Ferro e Manolo González, pola súa axuda e infinita paciencia comigo;

Paula Santalla, polas moitas horas de discusións enriquecedoras e estimulantes;

A miña familia, toda ela, pero principalmente meus pais, porque co seu esforzo abrironme a porta ao mundo da lingua; miñas dúas nais, miña nai e miña madriña, pola confianza e apoio que sempre me demostraron, por esixirme máis e, como non, por facer que viva coma unha raíña; meus irmáns –Rosa, Manu, Antonia, Julián–, pais en segunda instancia, sempre pendentes e dispostos para o que faga falta; Manuela, surtidora de lectura desde a infancia e niñeira consentidora;

A todos, citados polo nome ou implicitamente, grazas.

Índice:

Introdución.....	15
Capítulo 1. Determinación dun etiquetario	23
1.1 Introducción.....	23
1.2. A proposta de EAGLES	24
1.2.1. Clases de palabras.....	26
1.2.2. Lexicóns e corpus: etiquetarios distintos?	27
1.2.3. Atributos	29
1.2.3.1. Nome	29
1.2.3.2. Verbo	30
1.2.3.3. Adxectivo	31
1.2.3.4. Pronome e Determinante	33
1.2.3.5. Artigo.....	36
1.2.3.6. Adverbio	36
1.2.3.7. Adposición.....	37
1.2.3.8. Conxunción.....	38
1.2.3.9. Numeral	38
1.2.3.10. Única.....	39
1.2.3.11. Residual	40
1.3. O etiquetario de XIADA.....	41
1.3.1. Clases de palabras.....	41
1.3.2. Atributos	47
1.3.2.1. Substantivo (S)	47
1.3.2.2. Verbo (V).....	49
1.3.2.3. Adxectivo (A).....	55
1.3.2.4. Artigo (D)	58
1.3.2.5. Demostrativo (E)	59
1.3.2.6. Posesivo (M).....	60
1.3.2.7. Numeral (N).....	62
1.3.2.8. Interrogativo-exclamativo (G).....	68
1.3.2.9. Relativo (T).....	70
1.3.2.10. Indefinido (I)	71
1.3.2.11. Pronome (R)	75
1.3.2.12. Adverbio (W).....	88
1.3.2.13. Conxunción (C)	91
1.3.2.14. Locución (L).....	93
1.3.2.15. Preposición (P)	95
1.3.2.16. Interxección (Y).....	99
1.3.2.17. Sinais de puntuación (Q)	100

1.3.2.18. Categoría periférica (Z)	101
1.4. Datos	106
1.5. Outros etiquetarios para o galego	108
1.6. Conclusións	133

Capítulo 2. Elaboración da gramática formal do galego actual e a súa implementación nun lexicón..... 137

2.1. Introducción.....	137
2.2. O lexicón de XIADA.....	139
2.2.1. Estrutura xeral	139
2.2.2. Implementación do vocabulario no lexicón.....	144
2.2.2.1. Categorias invariables.....	145
2.2.2.2. Categorias variables: modelos formais.....	148
2.2.2.2.1. A flexión de xénero e número	148
2.2.2.2.2. A flexión de número	170
2.2.2.2.3. O verbo. Grupos formais	175
2.2.2.2.4. O verbo: a conjugación regular con enclíticos. Grupos formais ..	186
2.2.2.2.5. Verbos irregulares: modelos e raíces.....	193
2.2.2.3. Cambios na normativa oficial do 2003: repercusións no lexicón.....	206
2.3. Elementos problemáticos para un sistema computacional: as unidades multipalabra e as contraccións.....	212
2.3.1. Introducción	212
2.3.2. Unidades multipalabra e contraccións na proposta de EAGLES	215
2.3.3. Unidades multipalabra e contraccións noutros etiquetarios do galego.....	220
2.3.4. Unidades multipalabra e contraccións en XIADA	226
2.3.4.1. Unidades multipalabra.....	226
2.3.4.2. Contraccións	228
2.4. Un caso especial de unidades gramaticais complexas: formas verbais con enclíticos.....	247
2.4.1. EAGLES e sistemas do SLI-TECTRA e SLI-CTAG.....	247
2.4.2. Tratamento no sistema de XIADA	250
2.4.2.1. Xeneralidades	250
2.4.2.2. Lexicóns verbais: sublexicón <i>principal</i> e sublexicón <i>clíticos</i>	254
2.4.2.3. Secuencias de enclíticos válidas	257
2.4.2.4. Secuencias clíticas monosilábicas	259
2.4.2.5. Regras lingüísticas.....	263
2.4.2.5.1. Regras de clíticos.....	264
2.4.2.5.2. Regras de verbos con enclíticos.....	268
2.5. Os numerais multipalabra.....	273
2.6. Datos xerais	275
2.7. Datos verbais	278

2.8. Conclusións	284
Capítulo 3. A etiquetaxe automática	287
3.1. Introducción.....	287
3.2. O Etiquetador.....	289
3.2.1. Características.....	289
3.2.2. Rendemento e precisión	291
3.2.2.1. Tamaño do etiquetario	292
3.2.2.2. O corpus de adestramento	294
3.2.2.2.1. O núcleo de Xiada	295
3.2.2.2.2. O subcorpus xornalístico de ámbito temático específico	300
3.2.2.2.3. O subcorpus xornalístico xeral	310
3.2.2.3. O tratamento das palabras descoñecidas	313
3.2.2.4. O corpus de aplicación vs. o corpus de adestramento e o lexicón.....	314
3.3. A ambigüidade en galego: as anfiboloxías formais no corpus de adestramento.....	315
3.3.1. Tipos de ambigüidade.....	318
3.3.1.1. Homonimia intercategorial ou ambigüidade entre categorías.....	318
3.3.1.2. Homonimia intracategorial ou ambigüidade nos atributos.....	319
3.3.1.3. Homonimia categorial e atributiva plena.....	322
3.3.1.4. Homonimia segmental ou ambigüidade na identificación do <i>token</i>	327
3.3.2. Datos sobre a ambigüidade no corpus de adestramento	332
3.4. A ferramenta de desambiguación manual	334
3.5. Conclusións	354
Capítulo 4. A etiquetaxe manual.....	357
4.1. Introducción.....	357
4.2. O contexto na desambiguación manual	358
4.3. Características dos textos: erratas, variacións, dialectalismos, hipergaleguismos, castelanismos, etc.	360
4.3.1. O tratamento das erratas na etiquetaxe manual	362
4.3.2. O tratamento das desviacións da norma	369
4.3.2.1. Vacilacións ~ variacións no vocalismo	372
4.3.2.2. Vacilacións ~ variacións no consonantismo.....	380
4.3.2.2.1. Flutuacións B / V.....	380
4.3.2.2.2. Adverbios en <i>-mentes</i>	382
4.3.2.2.3. Simplificación de grupos consonánticos cultos.....	384
4.3.2.2.4. Flutuacións X / S	386
4.3.2.3. Vacilacións ~ variacións na acentuación.....	388
4.3.2.3.1. Acento diacrítico.....	388
4.3.2.3.2. Adverbios en <i>-mente</i>	392
4.3.2.3.3. Acentuación nas voces graves rematadas en ditongo crecente....	394

4.3.2.3.4. Acentuación nas voces agudas rematadas en ditongo decrecente	396
4.3.2.3.5. Acentuación nos hiatos	399
4.3.2.3.6. Acentuación esdrúxula na P4 e P5 do copretérito, pospretérito e antepretérito de indicativo e mais o pretérito de subxuntivo	403
4.3.2.4. Vacilacións ~ variacións na adscrición xenérica ou numérica	405
4.3.2.5. Vacilacións ~ variacións no caso acusativo e dativo	410
4.3.2.6. Vacilacións ~ variacións na adscrición categorial	420
4.4. Peculiaridades na etiquetaxe e lematización manuais de...	422
4.4.1. O pronome persoal tónico	422
4.4.2. <i>Un, unha, uns, unhas</i>	442
4.4.3. <i>Como</i>	445
4.4.4. <i>Que</i>	466
4.4.5. Categoría Verbo	487
4.4.5.1. Participio vs. adxectivo	487
4.4.5.2. Infinitivo vs. substantivo	493
4.4.5.3. Infinitivo vs. infinitivo conxugado	497
4.4.5.4. Verbos <i>poñer</i> e <i>pór</i> e derivados	498
4.4.6. <i>Claro, claro que...</i> : interxección?	499
4.4.7. Agrupacións con <i>outro</i> : ambigüidade no valor	500
4.4.8. <i>Propio</i> : indefinido ou adxectivo?	506
4.4.9. Combinacións tipo <i>boca arriba</i>	509
4.4.10. Combinacións tipo <i>anos despois</i>	512
4.4.11. Enderezos web, números de teléfono, identificadores, etc.	513
4.4.12. Acurtamentos	515
4.4.13. Soletreos e suspensións	518
4.4.14. Locucións	521
4.4.14.1. Nivel oracional vs. nivel supraoracional	521
4.4.14.2. Locucións conxuntivas	524
4.4.14.2.1. Preposición/locución prepositiva + <i>que</i> , locución conxuntiva? <i>Ata que, a pesar de que, para que, ó que</i>	524
4.4.14.2.2. Adverbio + <i>de</i> = locución prepositiva? + <i>que</i> , locución conxuntiva? <i>Antes de (que), despois de (que), logo de (que)</i>	528
4.4.14.2.3. <i>Antes que/ca, despois que/ca, logo que</i>	530
4.4.14.2.4. <i>Sempre que</i>	532
4.4.14.2.5. <i>Agora que, agora ben, ora que, ora ben</i>	535
4.4.14.2.6. <i>Cada vez que, unha vez que</i>	560
4.4.14.2.7. <i>Nin que, aínda que, mesmo que, por moito que, por máis que, por+adx+que</i>	572
4.4.14.3. Locucións adverbiais	588
4.4.14.3.1. <i>Pode que, poida que, pode ser que, poida ser que</i>	588
4.4.14.3.2. <i>Non obstante, sen embargo, con todo, así e todo, así a todo, en troques</i>	590
4.4.14.3.3. <i>Aínda así</i>	593
4.4.14.3.4. <i>Nada máis</i>	599
4.4.14.4. Locucións prepositivas	602
4.4.14.4.1. <i>Dentro de</i>	602
4.4.14.4.2. <i>Arredor de, ó redor de</i>	609

4.4.14.4.3. <i>Ademais de ~amais de, aparte de ~ á parte de</i>	610
4.4.14.4.4. <i>Debido a</i>	613
4.4.14.4.5. <i>Fóra de</i>	615
4.4.15. <i>Excepto, menos, quitado, quitando, sacado, sacando, salvo, senón:</i> problemas de categorización e tratamento en XIADA.....	616
4.4.15.1. Clasificación nos principais manuais actuais do galego.....	616
4.4.15.2. Breve revisión sobre os problemas para a delimitación categorial ...	618
4.4.15.3. O seu tratamento en XIADA	620
4.4.15.3.1. <i>Excepto, menos, salvo</i>	620
4.4.15.3.2. <i>Senón, senón que</i>	623
4.4.15.3.3. <i>Quitado/quitando, sacado/sacando</i>	632
4.5. Conclusións	639
Conclusións	641
APÉNDICES.....	647
Apéndice A. Etiquetario de XIADA.....	649
Apéndice B. Listaxe de etiquetas empregadas no CORGAetq	651
Apéndice C. Listaxe de etiquetas recollidas no kernel pero non documentadas na versión 2.1 do CORGAetq.....	671
Apéndice D. Listaxe de etiquetas usadas só no lexicón de Xiada.....	677
Apéndice E. Secuencias válidas de pronomes enclíticos en XIADA.....	681
Apéndice F. Ocorrencias dos verbos <i>informar</i> e <i>instar</i> na súa construción con complemento de réxime no CORGA	684
1) INFORMAR (A) ALGUÉN DE / SOBRE ALGO	685
a) no corpus de adestramento xornalístico	685
b) no corpus de adestramento de ficción	686
c) nas revistas do CORGA.....	686
d) nos xornais do CORGA	687
e) na ficción do CORGA	693
2) INSTAR (A) ALGUÉN A / PARA ALGO.....	703
a) no corpus de adestramento xornalístico	703
b) no corpus de adestramento de ficción	704
c) nas revistas do CORGA.....	704
d) nos xornais do CORGA	705
e) na ficción do CORGA	710
Referencias dos exemplos de prensa.....	713
Bibliografía.....	715
Referencias web	727

Introdución

A construción de recursos lingüísticos, entre os que se encontran os corpus ou bases de datos textuais, é necesaria en toda lingua para continuar profundando no seu coñecemento, mais tamén é fundamental para o procesamento da linguaxe natural.

Nos últimos anos, as novas tecnoloxías xorden como un parámetro máis de clasificación das linguas en función da súa presenza ou ausencia nelas e o galego non é alleo a esta corrente. Así, o *Plan xeral de normalización da lingua galega*¹ establece no último dos seus cinco obxectivos

Dotar o galego dos recursos lingüísticos e técnicos necesarios que o capaciten para vehicular a lingua moderna. [Xunta de Galicia, 2005: 39]

Ademais, as novas tecnoloxías destacan no PXNLG en forma de sector transversal sobre o que se estenden os obxectivos xerais:

Fomentar a presenza da lingua galega nas novas tecnoloxías.

Lograr unha oferta ampla e competitiva de produtos e recursos informáticos en galego.

Potenciar a presenza da lingua galega en internet.

Potenciar a investigación da tradución automática, o recoñecemento e a síntese de voz, e outras novas técnicas que faciliten a opción positiva no mercado da información e da comunicación, e que aseguren a libre circulación do galego nos sistemas avanzados da vida moderna. [Xunta de Galicia, 2005: 51]

Non obstante, o PXNLG non fai máis ca reflectir unhas necesidades das que xa eran conscientes os sectores implicados na defensa da lingua e sobre as que distintos grupos de investigación das tres universidades galegas, así como centros dependentes da Xunta de Galicia, levan anos traballando². Tanto é así, que a Secretaría Xeral de Política Lingüística, órgano superior da Administración autonómica en materia de promoción e ensino da lingua galega, a quen lle corresponde, entre outras, a función de programar e

¹ Xunta de Galicia (2005): *Plan xeral de normalización da lingua galega* (Santiago de Compostela).

² Os máis destacados, polo menos en resultados, localízanse no Instituto da Lingua Galega, no Seminario de Lingüística Computacional da Universidade de Vigo, no grupo COLE das Universidades da Coruña e Vigo e no Centro Ramón Piñeiro para a Investigación en Humanidades.

desenvolver medidas para o fomento do uso do idioma galego nos diferentes eidos sociais e económicos co fin de incrementar a dispoñibilidade de produtos e servizos en galego, sufraga e promove, a través do Centro Ramón Piñeiro para a Investigación en Humanidades –centro que inclúe entre os seus obxectivos fundacionais o deseño e desenvolvemento de recursos informáticos que axuden á incorporación da lingua galega ao mundo das novas tecnoloxías da información–, dous proxectos de investigación centrados en lingüística computacional aplicada ao galego. Son os denominados, respectivamente, *Corpus de Referencia do Galego Actual* (CORGA)³ e *Etiquetador/Lematizador do Galego Actual* (XIADA)⁴.

O primeiro deles persegue poñer accesible a través de internet un corpus documental representativo da lingua galega actual escrita que posibilite a obtención de datos para o estudo de aspectos morfolóxicos, sintácticos e léxicos da historia recente da nosa lingua.

Neste sentido, a forma máis rápida de dispoñer dunha grande cantidade de texto en formato electrónico á que se lle poidan aplicar técnicas de recuperación de información é con texto plano, sen ningún tipo de etiquetaxe; é dicir, sen que ás formas ortográficas contidas no corpus se lles asigne etiqueta ningunha. Constitúen mostras deste tipo o *Corpus de Referencia do Galego Actual* (CORGA), o *Corpus de Referencia del Español Actual* (CREA)⁵, o *Corpus de Referência do Português Contemporâneo* (CRPC)⁶ ou o *Corpus Técnico do Galego* (CTG)⁷, por citar só algúns. Este tipo de corpus, representativos da lingua común os tres primeiros e de carácter técnico o último, malia que sen etiquetaxe de ningún tipo, permiten xa entre outras posibilidades:

- documentar a vitalidade dunha palabra;
- coñecer cal é a forma máis frecuente nunha serie de sinónimos;
- achegarse ao significado dunha palabra grazas ao contexto;
- comprobar a presenza na lingua escrita de formas desaconselladas pola norma;
- comprobar a presenza de neoloxismos;

³ <http://corpus.cirp.es/corga>

⁴ <http://corpus.cirp.es/xiada>

⁵ <http://corpus.rae.es/creanet.html>

⁶ <http://alfclul.clul.ul.pt/CQPweb/crpc/index.php>

⁷ <http://sli.uvigo.es/CTG/>

- estudar o grao de asentamento da normativa oficial;
- achegarse á vitalidade e funcionamento de sufixos e prefixos;
- estudar o influxo (interferencias, préstamos...) doutras linguas.

Agora ben, todos eles presentan limitacións por canto non é posible abstraerse á rica morfoloxía que caracteriza as tres linguas peninsulares exemplificadas. É verdade que algunhas desas limitacións poden emendarse cun bo sistema de consultas no que sexa posible empregar comodíns –substitúen un ou máis caracteres– e mais operadores booleanos e de posición (*and, not, or, near...*) –permiten xogar con máis dunha forma gráfica na mesma consulta, favorecendo que afinemos máis a busca e, polo tanto, con menos ruído nos resultados–. Non obstante, en moitos casos, sobre todo ante grandes irregularidades léxicas, precisamos recorrer aos lemas e incluso á súa combinación cunha ou máis etiquetas morfosintácticas. É o caso, por exemplo, das consultas que abranguen algunha forma gráfica fortemente irregular ou combinacións categoriais fixas nas que varía un dos lemas, como pode ser o uso da perífrase *ir* + infinitivo, ou daquelas outras nas que por ambigüidade necesitamos poder contar con etiquetas que delimiten morfolóxica e categorialmente as distintas ocorrencias dunha forma gráfica, coma *a* preposición fronte a *a* artigo.

Conscientes das limitacións que impoñen as consultas por forma ortográfica, dende o *Centro Ramón Piñeiro para a Investigación en Humanidades* traballamos paralelamente na construción dun etiquetador e lematizador co fin de que se poidan facer consultas moito máis avanzadas e dar un salto cualitativo nas posibilidades de busca. É o fin último do segundo dos proxectos centrados en lingüística computacional que citamos anteriormente: a construción do *Etiquetador/Lematizador do Galego Actual* (XIADA), que pretende o desenvolvemento de ferramentas que permitan o recoñecemento e a análise automática do galego escrito actual. En especial, o desenvolvemento dun etiquetador e lematizador de alta precisión que permita etiquetar e lematizar automaticamente os documentos contidos no CORGA, de xeito que no sistema de consultas se poida empregar información referida non só a formas ortográficas senón tamén a lemas, formas gramaticais e/ou etiquetas morfosintácticas.

O punto de encontro natural de ambos os dous proxectos é a etiquetaxe automática de todos os documentos do CORGA, almacenados nun corpus análogo, só

que etiquetado, o *Corpus de Referencia do Galego Actual etiquetado automaticamente* (CORGAetqa), que gradualmente irá acollendo os documentos do CORGA para cuxo estilo se teña completado o adestramento do etiquetador. Naturalmente, este corpus será accesible mediante a consulta en liña a través dunha aplicación de recuperación e extracción de información que, cunha interface amigable, toma a consulta dun usuario sobre a grande base de datos textuais que, á fin e ao cabo, é calquera corpus, escolle entre todos os datos aqueles que responden ás condicións da busca realizada, e extrae os resultados trasladándoos a un formato predefinido para que a información resultante poida ser tratada con comodidade polo usuario⁸.

Eu, primeiramente como bolseira –dende 1995 ata 1999– e máis tarde en calidade de persoal contratado –dende o 2000 ata a actualidade–, formo parte de ambos os proxectos dende o seu inicio. Salvo a definición do etiquetario que, malia ser modificado lixeiramente en anos posteriores, foi obra de todo o equipo lingüístico que por volta do ano 1998 traballaba no proxecto⁹, a construción e mantemento da base de datos léxica e mais do corpus de adestramento que emprega o etiquetador XIADA, así como a elaboración de calquera outro recurso de tipo lingüístico que empregue o Etiquetador ou mellore a súa precisión, quero crer que con máis acertos ca erros, son responsabilidade miña.

Moitos son, xa que logo, os anos que levo traballando en lingüística de corpus, na recuperación e extracción de información e no recoñecemento e análise automática do galego. Froito desa experiencia nace o desexo de compendiar nun traballo global os distintos aspectos e recursos que, dende o punto de vista lingüístico, se precisan para construír un etiquetador e lematizador estatístico específico para o galego. É determinante tamén na decisión de describir o sistema XIADA o feito de que o galego sexa orfo neste ámbito, dado que na bibliografía da lingüística galega non existen máis ca un par de traballos moi breves que describen o etiquetario empregado nos corpus de galego que posúen algún tipo de anotación¹⁰.

Así mesmo, outros dous factores influíron de xeito determinante na consecución deste traballo. Por unha banda, o desexo de axudar a que o galego, lingua na que me

⁸ Proximamente estará dispoñible no enderezo <http://corpus.cirp.es/corgaetqa>

⁹ M^a Teresa Araújo García, M^a Ines Diz Gamallo, Eva M^a Domínguez Noya, Susana M^a García Rodríguez, M^a Teresa Monteagudo Cabaleiro, M^a Pilar Vázquez Grandes e Francisco García Gondar.

¹⁰ Véxase o Capítulo 1.

criei e na que vivo, figure entre as linguas que gozan de bos recursos computacionais; por outra, pensando no usuario que consulta o noso corpus etiquetado, dar a coñecer polo miúdo os múltiples problemas que crea traballar con lingua real e explicar as solucións que se adoptaron e mais os criterios que subxacen na etiquetaxe practicada.

Neste traballo descríbense os recursos lingüísticos que emprega o *Etiquedor/Lematizador do Galego Actual*, XIADA e desmiúzase a etiquetación realizada sobre o corpus de adestramento ou *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq)¹¹.

Así, o Capítulo 1 dá conta da determinación do etiquetario que se utiliza no sistema XIADA. Partindo das recomendacións do grupo EAGLES, explícase o establecemento de cada categoría gramatical e dos diferentes atributos que se consideran pertinentes para cada unha delas. Achéganse exemplos prácticos de etiquetaxe para as distintas clases gramaticais establecidas e, así mesmo, sempre en relación cos atributos, explícanse xa as peculiaridades que afectan á caracterización morfolóxica dalgúns ou de todos os elementos que se inclúen nunha categoría dada. Finalmente, realízase un estudo comparativo-contrastivo entre tres etiquetarios de galego: o empregado no *Corpus literario TECTRA inglés-galego*¹², orientado cara á tradución; o usado no *Corpus Técnico Anotado do Galego* (CTAG)¹³ de orientación terminolóxica especializada e, finalmente, o de XIADA, empregado no *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq), de marcada orientación cara á etiquetaxe de corpus e cara á recuperación de información para un amplo abano de usuarios potenciais.

No Capítulo 2 descríbese a formalización e implementación do dicionario léxico que lle serve ao etiquetador de recurso principal para identificar as formas ortográficas presentes nun texto e caracterizalas morfosintacticamente. Por unha banda, ao traballar cunha lingua con morfoloxía moi rica como é o galego, non é factible introducir nunha base de datos todas as formas flexionadas que se subsumen baixo un paradigma. Aquí explícase o método empregado para, sen reducir a variación flexiva, introducir no lexicón as 100.000 formas máis frecuentes do CORGA e as cerca de 50.000 entradas

¹¹ <http://corpus.cirp.es/corgaetq>

¹² <http://sli.uvigo.es/CLUVI/#tectraig>

¹³ <http://sli.uvigo.es/CTAG/index.html>

que posúe o *Vocabulario Ortográfico da Lingua Galega* (VOLGa) (González e Santamarina, 2004). Por outra banda, explícase tamén o novidoso tratamento aplicado sobre a grande cantidade de contraccións e as inxentes combinacións de forma verbal con segundas formas de artigo e/ou pronomes enclíticos que posúe o galego e que permite o recoñecemento e caracterización diferenciados de tantos constituíntes como conforman a unidade amalgamada. Por último, descríbese o sistema de regras lingüísticas que serven para axudarlle ao etiquetador tanto na segmentación e etiquetación de pronomes enclíticos como na segmentación e etiquetación das formas verbais cando van acompañadas de enclíticos.

No Capítulo 3 descríbese o etiquetador de XIADA e xustifícase a escolla do modelo estatístico. Logo, dáse conta, así mesmo, dos requisitos que esixe o tipo estatístico, ocupando un lugar destacado entre eles a construción e caracterización do corpus de adestramento ou *gold standard* que lle serve ao etiquetador de modelo para inferir a gramática da lingua coa que traballa. Finalmente, descríbese a actuación do etiquetador ante as palabras descoñecidas e inclúese unha breve tipoloxía das ambigüidades que potencialmente minguan a súa capacidade de acerto, mais ás que se ten que enfrontar.

O Capítulo 4 constitúe o groso deste traballo. Nel demóstranse as dificultades de traballar con lingua real, dificultades que se incrementan máis aínda por traballar cunha lingua coma o galego, cuxo proceso de normalización bate co castelán, lingua irmá, para o que aquí tratamos, en numerosas interferencias de tipo léxico e morfosintáctico. Nestas, céntrase a primeira parte do capítulo, onde se analizan as peculiaridades lingüísticas que caracterizan os textos sobre os que se executa o etiquetador e se expón o tratamento que reciben as erratas, as variacións gráficas, os castelanismos, hiperenxebrismos, etc. A segunda parte do capítulo constitúe, en termos xerais, unha descrición gramatical parcial do galego, pois nela refírense os criterios que guían a etiquetaxe realizada no CORGAetq. Ante a imposibilidade de dar conta neste traballo de absolutamente todos os detalles, preferiuse optar por tratar aqueles aspectos referidos á etiquetaxe e lematización de:

- elementos que en xeral non se inclúen nos manuais de consulta: enderezos, acurtamentos, identificadores, etc.;
- unidades nas que a ambigüidade é moi alta: *que, como*;

- elementos pertencentes a categorías moi próximas: adxectivo fronte a participio ou substantivo fronte a infinitivo;
- formas comúns a dous lemas distintos: *pór* vs. *poñer* e derivados;
- combinacións que reciben unha caracterización *ad hoc* para facilitar a recuperación de información: tipo *boca arriba* ou tipo *anos despois*.

Ocupa gran parte deste último capítulo a delimitación e caracterización dos distintos tipos de locucións recoñecidos no sistema: as conxuntivas, as preposicionais e as adverbiais, para as que se asentán os criterios xerais empregados no seu recoñecemento. En cada un dos tres apartados recóllese, ademais, a análise aplicada a aquelas que, tendo unha frecuencia de uso importante, consideramos merecentes dunha explicación, ben porque serven de exemplo para o establecemento da súa subclase, ben porque reciben unha análise que en principio non esperan os usuarios.

Finalmente, a análise que se realiza ao final do capítulo das partículas exceptivas *salvo*, *incluso*, *menos* e *senón* exemplifica como a lingüística de corpus pode inverter a metodoloxía utilizada ata hai pouco nas descrições lingüísticas, proporcionando unha casuística real e completa que axude a profundar no coñecemento da lingua para, partindo dos exemplos, chegar á teoría.

Así pois, e a modo de resumo, este traballo dá conta da determinación do conxunto de etiquetas que se empregan no sistema XIADA para anotar morfoloxicamente un documento escrito en galego (Capítulo 1); describe a estrutura interna do lexicón onde se almacena, basicamente, a información de etiqueta e lema que lle corresponde a cada unidade léxica (Capítulo 2) e recolle a construción dun corpus de adestramento como requisito imprescindible para a aplicación dun etiquetador estatístico-probabilístico (Capítulo 3) e mais os criterios de lematización e etiquetaxe que, a grandes trazos, caracterizan a gramática que subxace nos documentos etiquetados e que vai ser inferida polo etiquetador de XIADA (Capítulo 4).

Cada capítulo inclúe como última epígrafe as conclusións máis destacadas que se derivan da reflexión sobre a súa lectura ou elaboración, mentres que na conclusión final realízase unha valoración sobre o conxunto do traballo, destacando os logros obtidos e enumerando as parcelas que requiren máis atención para incrementar a taxa de acerto da etiquetaxe final.

Capítulo 1

Determinación dun etiquetario

Antes de proceder a anotar morfoloxicamente un documento hai diversos aspectos que considerar. En primeiro lugar débese determinar o conxunto de etiquetas que se van aplicar. En segundo lugar, é preciso deseñar e construír o lexicón onde se almacenará, como mínimo, a información de etiqueta e lema para cada unidade léxica e, en terceiro lugar, débense especificar os criterios que determinarán que etiqueta se lle debe aplicar a cada elemento gramatical.

Neste capítulo imos tratar a determinación do conxunto de etiquetas para o galego destinadas a codificar o lexicón e a etiquetar, posteriormente, corpus.

1.1 Introducción

A relación de etiquetas empregadas para anotar morfoloxicamente un texto denomínase etiquetario ou *tag-set*. En relación coa profundidade de análise que se pretenda alcanzar, este conterá un número maior ou menor de etiquetas. É primordial, polo tanto, antes de determinar o sistema de etiquetación, decidir que obxectivo se persegue coa anotación e, en consecuencia, a que nivel de análise se quere chegar. Debe terse en conta tamén o modo en que se vai realizar a anotación, pois se esta vai ser semiautomática ou automática hai que ver de conxugar os intereses lingüísticos cos límites que impón un etiquetador automático.

O noso obxectivo é analizar morfoloxicamente e dun xeito automático calquera texto do galego actual coa finalidade de lles proporcionar aos usuarios do *Corpus de Referencia do Galego Actual (CORGA)*¹⁴ a posibilidade de acceder a material lingüístico real, de forma rápida e eficiente, realizando consultas, entre outras moitas opcións de busca, por etiqueta e lema. Por este motivo, na construción do *tag-set* decidimos primar a descrición morfolóxica e prescindir, case, da sintáctica.

¹⁴ Véxase <http://corpus.cirp.es/corga>

O etiquetario para o galego foi desenvolvido no marco do proxecto *Construcción, etiquetación e lematización do Corpus de Referencia do Galego Actual*¹⁵ como unha ferramenta para a análise computacional e a anotación automática dos textos do CORGA.

Para a elaboración desta etiquetaxe seguimos as directrices de EAGLES (*Expert Advisory Group on Language Engineering Standards*)¹⁶ e os textos básicos da lingüística galega existentes na última década do século XX. A *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986) serviunos de soporte teórico principal, mais tamén nos apoiámos na *Nova Gramática para a aprendizaxe da lingua* (Costa Casas et al., 1988).

A escaseza de traballos gramaticais xerais para o galego foi suplida con estudos, sobre todo, do castelán e, en menor medida, do portugués: a *Gramática de la lengua española* (Alarcos, 1994), o *Esbozo de una nueva gramática de la lengua española* (RAE, 1973), o estudo *Las categorías gramaticales. Relaciones y diferencias* (Bosque, 1991) ou a *Nova Gramática do Portugués Contemporáneo* (Cunha e Cintra, 1984) axudaron na delimitación do etiquetario e na clasificación posterior do léxico constitutivo.

Tivemos tamén en consideración o camiño xa percorrido na etiquetación de corpus por linguas coma o castelán (Sánchez, 1997), catalán (Morel et al., 1998) ou portugués (Reis e Dias, 1998).

1.2. A proposta de EAGLES

O grupo EAGLES encargouse de avaliar recursos lingüísticos existentes para diferentes linguas europeas¹⁷ co fin de acadar unha estandarización na codificación do léxico e a etiquetación de corpus. Neste sentido, EAGLES non propón un etiquetario concreto senón que máis ben apunta as directrices que se han de ter en conta á hora de elaborar un. Destacan nas súas recomendacións os principios de flexibilidade, harmonización e exportabilidade.

¹⁵ Véxase <http://corpus.cirp.es/xiada>

¹⁶ Leech e Wilson, 1994 e 1996 e Monachini e Calzolari, 1996.

¹⁷ Compararon os sistemas de codificación dos lexicóns de MULTILEX e GENELEX, a aplicación AlethDic de GENELEX, as especificacións do proxecto NERC relativas a corpus e mais as propostas de Leech e Wilson.

A proposta de EAGLES presenta unha estrutura xerárquica na que se distribúe a información lingüística en tres niveis. Isto permítelle ao usuario escoller o nivel de análise e proporcionalle flexibilidade á proposta. Na súa descrición, EAGLES asocia cada nivel cun grao de profundación de análise ou obrigatoriedade. Distingue deste xeito entre nivel obrigatorio (L0), nivel recomendable (L1) e nivel opcional (L2).

O primeiro nivel de análise, o denominado L0, é considerado imprescindible e nel as unidades léxicas deben ser clasificadas nas principais clases de palabras ou categorías morfolóxicas:

Only one attribute is considered obligatory: that of the major word categories, or *parts of speech*. [Leech e Wilson, 1996]

No segundo nivel do estándar de EAGLES (L1) recoméndase a aplicación de atributos gramaticais que permitan o recoñecemento morfolóxico pleno de calquera unidade léxica:

Level 1 (EAG-L1) presents the grammatical features, i.e. the agreement features such as Gender, Number, Person, etc., which are usually encoded in lexicons and corpora: these are considered as recommended features (in Leech & Wilson these are defined as optional) constituting the minimal common core set of features for the PoS. [Monachini e Calzolari, 1996]

Finalmente, o terceiro nivel de análise proposto por EAGLES é totalmente opcional e nel poden codificarse:

- 1) (L2a) atributos ou valores xenéricos, aplicables á maioría das linguas estudadas, os cales normalmente non se codifican nun corpus por proporcionar información que excede o nivel morfosintáctico. A información semántica que achega a calidade de *countable* ou *non countable* proposta para a categoría Nome constitúe un exemplo de atributo opcional xenérico.
- 2) (L2b) atributos ou valores específicos dunha lingua ou de varias, pero non aplicables á maioría de linguas europeas.

Dado o obxectivo que nos propuxemos e a súa finalidade, consideramos que debiamos obter a caracterización morfolóxica completa de calquera unidade léxica, razón pola que incorporamos os tres niveis no noso etiquetario. Prescindimos en xeral, iso si, das distincións semánticas incluídas no L2a e non establecemos diferenciacións

de ningún tipo entre o L1 e o L2 no deseño do *tag-set*, onde os distintos valores atributivos figuran ao mesmo nivel. Por exemplo, na etiqueta *Scma* correspondente á análise dunha unidade léxica coma *paraugas*:

“S” representa a clase de palabra, substantivo;

“c” proporciona información sobre o *tipo*, común;

“m” indícanos que o seu xénero é masculino e

“a” sinala que pode ser *singular* ou *plural*; é dicir, apunta a imposibilidade de decantarnos por un dos valores do atributo *número* a non ser que a concordancia con algún elemento do contexto nolo aclare.

1.2.1. Clases de palabras

As grandes clases de palabras que propoñen Leech e Wilson¹⁸ (1996) son as seguintes:

1.	N [noun]	2.	V [verb]	3.	AJ [adjective]
4.	PD [pronoun/determiner]	5.	AT [article]	6.	AV [adverb]
7.	AP [adposition]	8.	C [conjunction]	9.	NU [numeral]
10.	Y [interjection]	11.	U [unique/unassigned]	12.	R [residual]
13.	PU [punctuation]				

Meses máis tarde, Monachini e Calzolari¹⁹ cuestiónanse a agrupación de Pronomes e Determinantes nunha única categoría, por estar esta demasiado orientada a facilitar a anotación de corpus, e deciden delimitalas como clases independentes:

(...) the previous merging of the two categories (at least at L1, with the possibility of splitting the two categories at a more fine-grained level) seemed, in the first instance, to be the best solution to cope with the requirements of many corpus practices, that keep the two categories undistinguished, and the best attempt to reconcile lexicon specifications and corpus tagsets. This choice, however, was eventually felt to be too corpus-oriented and the MULTTEXT partners expressed their opinion in favour of having, at the lexicon level, two different categories for Pronouns and Determiners. [Monachini e Calzolari, 1996]

¹⁸ Marzo de 1996: *Recommendations for the Morphosyntactic Annotation of Corpora*. Disponible en <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>

¹⁹ Agosto de 1996: *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons end Corpora. A common Proposal and Applications to European Languages*. Disponible en <http://www.ilc.cnr.it/EAGLES96/morphsyn/morphsyn.html>

Así, na nova proposta figuran as seguintes categorías:

1.	Noun	2.	Verb	3.	Adjective
4.	Pronoun	5.	Determiner	6.	Article
7.	Adverb	8.	Adposition	9.	Conjunction
10.	Numeral	11.	Interjection	12.	Unique
13.	Residual				

Se nos fixamos nos números, esta táboa contén a mesma cifra de clases gramaticais delimitadas ca na proposta primaria. A diferenza estriba en que nesta última clasificación non inclúen os signos de puntuación, aínda que si recoñecen a súa necesidade na etiquetación de corpus:

Sometimes tag classes are in reality different from lexical descriptions. For example, classes for punctuation are needed and certain types of semantic or pragmatic or lexical information can be present in the tags (e.g. the days of the week). [Monachini e Calzolari, 1996]

1.2.2. Lexicóns e corpus: etiquetarios distintos?

A causa que motiva a división da clase Pronome/Determinante en dúas clases diferenciadas lévanos a cuestionar teoricamente a adopción de dous conxuntos de etiquetas distintos, un destinado a codificar o léxico e outro reservado para etiquetar o corpus.

EAGLES, na súa proposta, recomenda aplicar a distinción entre Pronomes e Determinantes só ao etiquetario do lexicón e non ao etiquetario que se empregue para anotar e desambiguar corpus, porque os obxectivos que se poden acadar con un e con outro son diferentes. Así, co lexicón aspírase a describir pormenorizadamente a morfosintaxe xeral dunha lingua mentres que co *tag-set* de corpus búscase a elaboración dun corpus etiquetado morfoloxicamente por medio de desambiguación automática:

Lexical descriptions, it was recognised, should aim, indeed, at a general and fine-grained description of the language which is independent of particular applications, while, given a set of practical reasons -- state-of-the-art tagging techniques and computability (see Bel et al. (1995)) -- broader categories are to be preferred for the tagsets where collapsing decisions are to be made. [Monachini e Calzolari, 1996]

Segundo se desprende da súa proposta, aínda que o ideal sería poder aplicar a codificación do lexicón á anotación de corpus, resultaría moi difícil conseguir unha alta taxa de precisión cun etiquetador automático e, ademais, o corpus de adestramento

destinado a capacitar o etiquetador tería que ser de gran tamaño para poder documentar co-ocorrencias infrecuentes.

No seu favor esgrimen as prácticas levadas a cabo cos etiquetarios do francés no proxecto MULTEXT e a experiencia, entre outros, do Penn Treebank²⁰ para o inglés, como argumento definitivo da amálgama e consecuente redución de etiquetas empregadas para anotar corpus:

For example, in the proposal for French presented in the MULTEXT document (Bel et al., 1995), there are 249 different lexical descriptions, but only 74 collapsed corpus tags. Experience (Church, UPenn Treebank, IBM France, etc.) shows that the tagset should be under 100 tags. In fact, the Penn Treebank project collapsed many tags compared to the original Brown tagset, and got better results. [Monachini e Calzolari, 1996]

Evidentemente, ante un etiquetario reducido, as probabilidades de éxito dun etiquetador automático incrementáanse de xeito notable, pero iso non implica que empregando un *tag-set* amplo non poidan lograrse uns resultados máis que aceptables.

Nós propuxémonos, na etapa inicial, non tomar como referencia este propósito de manter etiquetarios separados para o lexicón e para a etiquetación de corpus. Decidimos que a todas as entradas do lexicón se lles proporcionase a información morfolóxica plena que as caracterizase, con todos os atributos e valores pertinentes para esta función. Posteriormente, cando chegase o momento de etiquetar corpus, se os resultados conseguidos estivesen moi lonxe dos desexados, poderíamos anular algúns atributos ou incluso algunhas distincións categoriais. O importante era crear un *tag-set* pormenorizado, sinxelo na súa estrutura, manexable e adaptable a outros sistemas ou aplicacións. Co paso do tempo e o avance no proxecto das tarefas de lematización e etiquetación confirmouse a nosa visión.

Actualmente, as porcentaxes de acerto que presentan os etiquetadores existentes para unha lingua como o castelán ou o inglés son do 95%/97% (Graña, 2000: 137-165) e, segundo Jurafsky (Jurafsky, 2000: 296-297), os etiquetarios máis usados para o inglés son o *Penn Treebank* que consta de 45 etiquetas e é o que se usou no corpus *Brown*²¹; o *C5* de 61 etiquetas empregado polo programa *CLAWS*²² para etiquetar o *British*

²⁰ Véxase <http://www.cis.upenn.edu/~treebank>

²¹ http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html

²² <http://ucrel.lancs.ac.uk/claws/>

Nacional Corpus (BNC)²³ e o C7, de 146 etiquetas, utilizado para etiquetar un subconxunto do BNC.

Nós demostramos que cun etiquetario bastante máis grande ca o empregado pola maior parte de etiquetadores –arredor de 400 etiquetas distintas– acadamos un acerto do 95,99%²⁴, situándonos en niveis semellantes aos existentes para o inglés ou o castelán.

1.2.3. Atributos

Neste apartado reproducimos a proposta de EAGLES no concernente aos atributos e valores pertinentes para cada unha das clases de palabras por eles delimitadas. A maiores, daremos conta do modelo seguido polo castelán e polo portugués nas aplicacións de EUROTRA e GENELEX, cando este non coincida coa proposta presentada. Se non se realiza ningunha aclaración sobre a lingua a que corresponde entenderase que son atributos e/ou valores comúns ás dúas.

1.2.3.1. Nome

Para o Substantivo propoñen os seguintes atributos:

L0	Substantivo
L1	Tipo: común, propio
	Xénero: masculino, feminino
	Número: singular, plural
L2	Xénero: común
	Número: invariable

Dende EAGLES, aínda que nas aplicacións dos proxectos castelán e portugués por eles estudados non se levou á práctica, recomendan incluír na caracterización do substantivo a calidade de *contable* ou *non contable*, posto que consideran que este criterio semántico é pertinente para a clasificación do Nome nas dúas linguas.

Na aplicación EUROTRA para o castelán, a maiores, empregan no xénero un novo valor (*transart*) para, nos substantivos femininos que comezan por “a” tónico, discriminar o alomorfo do artigo feminino (*el área*) do artigo masculino (*el niño*). Desmárcanse, con todo, da proposta de EAGLES e das demais aplicacións recollidas

²³ <http://www.natcorp.ox.ac.uk>

²⁴ Domínguez et al., 2009.

para outras linguas ao integrar os pronomes na categoría Nome. Así, para distinguir os substantivos dos pronomes recorren ao atributo *nform* que ten os seguintes valores:

Atributo	Valor	Exemplo	Etiqueta
nform	<i>normal</i>	casa	norm
	<i>clítico</i>	le	cli
	<i>pronome</i>	él	pro

A introdución dos pronomes na mesma clase ca os substantivos obrigaos a incluír atributos específicos dos distintos tipos de pronomes considerados como son a *persoa* ou o *caso*, pertinentes por exemplo para o pronome persoal, ou exclusivos dun subtipo como é o *número do posuidor* para o posesivo, ao que, como veremos cando cheguemos á categoría Pronome, tamén consideran un pronome.

1.2.3.2. Verbo

EAGLES propón os atributos seguintes para caracterizar o Verbo:

L0	Verbo
L1	Tipo: principal, auxiliar, modal
	Forma verbal: finita, non finita
	Modo: indicativo, subxuntivo, imperativo, condicional, infinitivo, participio, xerundio
	Tempo: presente, imperfecto, futuro, pasado
	Persoa: 1, 2, 3
	Número: singular, plural
L2	Xénero: masculino, feminino
	Tipo: copulativo, semiauxiliar
	Aspecto: perfectivo, imperfectivo
	Voz: activa, pasiva
	Reflexividade: reflexivo, non reflexivo
	Tipo de verbo principal: transitivo, intransitivo, impersoal
	Clíticos: clíticos, non clíticos

Á hora de aplicar este conxunto de atributos ás linguas específicas máis próximas á nosa, atopámonos con que, tanto no modelo EUROTRA, para o castelán, como no GENELEX, para o portugués, se prescindiu de calquera delimitación sintáctica ou semántica, limitándose a realizar unha caracterización estritamente morfolóxica.

Nas táboas que seguen recolleemos os atributos verbais aplicados ás dúas linguas:

	Forma verbal	Modo	Tempo	Persoa	Número	Xénero
	finita	indicativo	presente	primeira	singular	masculino
	non finita	subxuntivo	pasado	segunda	plural	feminino
Castelán		imperativo	futuro	terceira		
		infinitivo	condicional			
		xerundio	imperativo ²⁵			
		participio				

	Modo	Tempo	Persoa deíctica	Persoa concordancia	Xénero	Número
	indicativo	presente	1	1	masculino	singular
	subxuntivo	imperfecto	2	2	feminino	plural
Portugués	imperativo	perfecto	3	3		
	condicional	futuro				
	infinitivo	pasado				
	participio	pluscuamperfecto				
	xerundio					

1.2.3.3. Adxectivo

Os atributos propostos para caracterizar os adxectivos son:

L0	Adxectivo
L1	Tipo: cualificativo, posesivo, indefinido, cardinal, ordinal
	Grao: positivo, comparativo, superlativo
	Xénero: masculino, feminino
	Número: singular, plural
	Número do posuidor: singular, plural
L2	Persoa: 1, 2, 3
	Uso: atributivo, predicativo
	Tipo de modificación: premodificador, posmodificador
	Xénero: común
	Número: invariable

²⁵ Aparece recollido tamén como valor temporal pero cremos que se trata dun erro porque non se explica en ningún momento a súa inclusión dentro dos distintos tempos verbais, explicación que si aparece para o condicional.

Na proposta inicial de EAGLES (Leech e Wilson, 1996), curiosamente, non aparece o atributo *tipo* para a categoría Adxectivo; este xorde en Monachini e Calzolari (1996) para facer posible a marcación das tres funcións ou posicións que os tradicionalmente denominados adxectivos determinativos poden desempeñar ou ocupar, seguindo o modelo xa establecido por GENELEX:

Moreover, the GENELEX model distinguishes 3 possible functions, namely:

- 1 Adjective: "Le chien est nôtre."
- 2 Determiner: "nôtre chien."
- 3 Pronoun: "le nôtre." [Monachini e Calzolari, 1996]

O establecemento do atributo *tipo* en Monachini e Calzolari parece, cando menos, contraditorio. Si é certo que:

The GENELEX and AlethDic models reflect a distinction which is also made in traditional grammars of other Romance languages (e.g. Italian), i.e. a first disjunction between *qualificatif* and *indicatif* Adjectives. The latter are those Adjectives which also have pronominal function, and in different grammatical traditions are called Determiners in their adjectival function. The former are all the other Adjectives.

Indicative Adjectives are further divided into *possessives*, *demonstratives*, *relatives*, *indefinites*, *numerals* (ordinals and cardinals), *interrogatives* and *exclamatories*. [Monachini e Calzolari, 1996]

pero os demostrativos, exclamativos, interrogativos e relativos brillan pola súa ausencia como subtipos do Adxectivo. Non estamos, pois, ante a distinción tradicional *strictu sensu* entre adxectivos cualificativos e determinativos. Logo, como se verá no Numeral, a pesar de preconizar a súa independencia como categoría, inclúen na súa propia proposta os dous subtipos de numerais: cardinais e ordinais. Finalmente, recoñecen a inclusión dos posesivos, indefinidos e os dous tipos de numerais polo desiderátum da *majority of the partners*.

Na aplicación ao castelán, na nosa opinión acertadamente, non se segue a proposta de EAGLES e os atributos considerados son os pertinentes para os chamados adxectivos cualificativos:

Grao	Xénero	Número	Posición
positivo	masculino	singular	premodificador
comparativo	feminino	plural	posmodificador
superlativo	común	invariante	indiferente

O portugués prescinde dos atributos *uso* e *tipo de modificación* de EAGLES e, igual que fixera na categoría verbal, introduce, a maiores, a diferenciación entre a persoa semántica e a que marca a concordancia:

Tipo	Grao	Xénero	Número	Persoa deíctica	Persoa concordancia	Número posuidor
cualificativo	positivo	masculino	singular	1	1	singular
posesivo	comparativo	feminino	plural	2	2	plural
indefinido	superlativo			3	3	
cardinal						
ordinal						

1.2.3.4. Pronome e Determinante

Imos tratar na mesma epígrafe a caracterización das clases Pronome e Determinante, recollida en Monachini e Calzolari (1996), para que se advirtan máis facilmente as similitudes e diferenzas existentes entre elas.

Na proposta de EAGLES, na categoría Pronome intégrase o persoal e os denominados determinativos en función pronominal; a categoría Determinante, pola súa banda, acolle os tradicionalmente chamados determinativos en función adxectivo²⁶:

²⁶ Tomamos literalmente de EAGLES os termos “función pronominal” e “función adxectivo” sendo conscientes de que dende o punto de vista da sintaxe funcional, en realidade, mesturan función e unidade, pois pronome e adxectivo son unidades que poden desempeñar diferentes funcións sintácticas e, sintacticamente, non existe ningunha “función pronominal” nin “función adxectivo”. Non obstante, o valor que aquí se lle outorga é dun xeito claro, respectivamente, o do seu funcionamento como NÚCLEO e como DETERMINANTE.

L0	Pronome	L0	Determinante
L1	Tipo: demostrativo, indefinido, posesivo, interrogativo, relativo, persoal, reflexivo, recíproco	L1	Tipo: demostrativo, indefinido, posesivo, interrogativo, relativo
	Persoa: 1, 2, 3		Persoa: 1, 2, 3
	Xénero: masculino, feminino, neutro		Xénero: masculino, feminino, neutro
	Número: singular, plural		Número: singular, plural
	Número do posuidor: singular, plural		Número do posuidor: singular, plural
	Caso: nominativo, xenitivo, dativo, acusativo, oblicuo, obxecto		
L2	Xénero: común	L2	Xénero: común
	Número: invariable		Número: invariable
	Cortesía: cortés, non cortés		
	Tonicidade: átono, tónico		

Deixando a un lado a conveniencia, ou non, de discriminar entre pronomes *persoais* por unha banda e *reflexivos* e *recíprocos* por outra, se extraemos da táboa os atributos específicos do subtipo *persoal*, isto é, *caso*, *cortesía* e *tonicidade*²⁷, atopámonos con que os atributos e os valores establecidos para a clase Pronome coinciden exactamente cos propostos para a clase Determinante. Reflictese deste xeito a categorización dos chamados determinativos priorizando a súa función e non a súa clase. A pregunta que xorde e cuxa resposta sería interesante coñecer é por que non se adoptou a mesma solución ca a proposta para a clase Numeral: un atributo co que se especifica a función do elemento? Porén, non atopamos ningunha explicación ao respecto en Monachini e Calzolari (1996).

A aplicación ao portugués presenta no Pronome dúas diferenzas de vulto con respecto á proposta de EAGLES. A primeira dáse na caracterización da *cortesía*. Como xa vén sendo habitual cando está presente o atributo *persoa*, o portugués distingue entre a persoa que esixe a concordancia e a que lle corresponde deicticamente. Non podemos, polo tanto, negar o recoñecemento das formas corteses na aplicación portuguesa, só que esta non emprega o atributo *cortesía* nin o fai unicamente no pronome persoal, senón sempre que estea implicada a persoa: verbo, posesivo e pronome persoal. A segunda

²⁷ Permitímonos a licenza de modificar a orde destes tres atributos para que a comparación entre as dúas categorías resulte máis cómoda. No orixinal aparece o *caso* antes do *número do posuidor* (L1) e a *cortesía* e a *tonicidade* como primeiros atributos xenéricos opcionais (L2).

diferenza xorde porque non se organizan as formas do pronome persoal nas súas dúas series características: tónica e átona.

Na clase Determinante as diferenzas acentúanse. A aplicación portuguesa engade no *tipo* os valores *definido* e *indefinido*, destinados a identificar o artigo que para EAGLES constitúe unha categoría propia, e o *cardinal*. Logo, ademais de non ter en conta os valores específicos opcionais para o xénero e o número, prescinden inexplicablemente dos atributos *persoa*²⁸ e *número do posuidor*.

Na aplicación EUROTRA para o castelán, como xa explicamos na clase Nome, os pronomes integráronse cos substantivos sendo o atributo *nform* o que os clasifica en *normal*, *pronome* e *clítico*. A maiores, para axudar na caracterización dos pronomes, empregaron dous atributos máis:

Dtype, which helps to distinguish between possessive (*poss*) and non-possessive (*non-poss*) pronouns.

Whmor, used to distinguish relative pronouns (valued as *rel*) and interrogative pronouns (valued as *int*) from the rest of the nominals (valued *none*). [Monachini e Calzolari, 1996]

En resumo, na clase Nome na aplicación de EUROTRA, con estes tres atributos distinguen entre (Monachini e Calzolari, 1996):

Sp. example	Nform	Dtype	Whmor	Class
casa	<i>normal</i>	<i>non-poss</i>	<i>none</i>	normal noun
yo	<i>pro</i>	<i>non-poss</i>	<i>none</i>	personal pro
este	<i>pro</i>	<i>non-poss</i>	<i>none</i>	demonstrative pro
algún	<i>pro</i>	<i>non-poss</i>	<i>none</i>	indefinite pro
mi	<i>pro</i>	<i>poss</i>	<i>none</i>	possessive pro
mío	<i>pro</i>	<i>poss</i>	<i>none</i>	possessive pro
cuyo	<i>pro</i>	<i>poss</i>	<i>rel</i>	relative pro
quien	<i>pro</i>	<i>non-poss</i>	<i>rel</i>	relative pro
quién	<i>pro</i>	<i>non-poss</i>	<i>int</i>	interrogative pro
me	<i>cli</i>	<i>non-poss</i>	<i>none</i>	clitic

Na delimitación que se leva a cabo entre os distintos tipos de pronome e de determinantes na aplicación ao castelán non está claro se é a que se realiza en EUROTRA ou se se trata da que realiza EAGLES para esta lingua. Os datos inducen a

²⁸ A pesar de haber posesivos implicados, na clase Determinante non se teñen en conta os atributos *persoa* e *número do posuidor*.

pensar que se trata da adaptación ao castelán de EAGLES. Con todo, independentemente da súa autoría, os subtipos de Pronome e Determinante, así como os atributos e valores aplicados nas dúas clases, son coincidentes na súa totalidade coa proposta de EAGLES, salvo nos trazos específicos do castelán, como poden ser:

- 1) ausencia do atributo xénero e do valor *invariable* no número no pronome relativo (*que, quien*) fronte á presenza dos valores *masculino* e *feminino* e *singular* e *plural* no determinante relativo (*cuyo/s, cuya/s*);
- 2) presenza unicamente do valor xenérico *neutro* no pronome demostrativo;
- 3) discriminación no *caso* dos valores *nominativo, acusativo, dativo* e *oblicuo*.

1.2.3.5. Artigo

Os atributos pertinentes para describir a clase Artigo, segundo EAGLES, son os seguintes:

L0	Artigo
L1	Tipo: definido, indefinido
	Xénero: masculino, feminino, neutro
	Número: singular, plural

Na aplicación ao portugués, fronte á castelá, non se recoñece unha clase Artigo, senón que os valores *definido* e *indefinido* figuran como dous tipos diferentes da clase Determinante, ao lado de *demostrativo, posesivo, cuantificador*, etc.

1.2.3.6. Adverbio

Admitindo que son numerosas as posibles subclasificacións dos adverbios fundamentándoas en criterios semánticos e sintácticos, a proposta de EAGLES para caracterizalos é a seguinte:

L0	Adverbio
L1	Tipo: xeral, partícula
	Grao: positivo, comparativo, superlativo
L2	Polaridade: interrogativo/relativo, non interrogativo/relativo

Chama a atención nesta clase que en Leech e Wilson (1996) o valor *partícula* do atributo *tipo*, aplicado especialmente no inglés para adverbios coma *out*, *off* ou *up*, figure entre os opcionais, mentres que se establece entre os recomendados en Monachini e Calzolari:

Another interesting fact is the inclusion of Particles among the values of the feature **Type** by the Leech & Wilson proposal. The treatment of Particles appears to be somewhat complicated: NERC suggests their inclusion in the preposition category; in GENELEX and MULTILEX they form a separate category without any attribute.

In the present EAGLES proposal, the values proposed for the feature **Type** at Level 1 (recommended) reflect a first distinction between general adverbs and particles. [Monachini e Calzolari, 1996]

Co atributo *polaridade* distínguense no adverbio as formas interrogativas-exclamativas ou relativas (en inglés *when*, *where*, *how*, *why*) das demais; os valores *comparativo* ou *superlativo* do atributo *grao* permiten diferenciar as formas sintéticas, comparativas ou superlativas, características dun número reducido de elementos adverbiais fronte ao valor *positivo* que se aplica por defecto para o resto dos casos.

Na aplicación ao portugués non diferencian *tipo* e no grao caracterizan só os adverbios *mais*, *menos* e *tão* como indicadores da comparación de superioridade, inferioridade e igualdade, respectivamente.

No modelo EUROTRA para o castelán empregan o atributo *whmor* para discriminar entre os adverbios interrogativos, relativos e os demais. A maiores, no nivel sintáctico, realizan unha clasificación semántica do adverbio (*tempo*, *modo*, *lugar*, etc.) para tratar fenómenos relacionados coa adición e aglomeración de adverbios.

1.2.3.7. Adposición

Os atributos propostos para caracterizar esta clase son:

L0	Adposición
L1	Tipo: preposición, posposición, circumposición
	Formación: simple, contracción
	Xénero: masculino, feminino, neutro
	Número: singular, plural

O portugués non reconece no *tipo* máis ca a preposición e, seguindo o modelo multilingüe de GENELEX, as contraccións de preposición e artigo trátanas como unha clase especial de unidade morfolóxica.

Na aplicación ao castelán, no nivel morfolóxico, non se considera ningún atributo e as contraccións *al* e *del* clasifícanas nunha categoría creada *ad hoc*.

1.2.3.8. Conxunción

A proposta de EAGLES para a clase Conxunción é a seguinte:

L0	Conxunción
L1	Tipo: coordinante, subordinante
L2	Subtipo coordinante: simple, correlativa, inicial, non inicial
	Subtipo subordinante: con forma verbal finita, con forma verbal non finita, introduce comparación

En Monachini e Calzolari destácase o consenso multilingüe acadado para a Conxunción, pero este límitase ao recoñecemento da categoría e do atributo *tipo*:

A large core of agreement emerges as regards Conjunctions. All the analysed systems agree as to the distinction between coordinating and subordinating Conjunctions. The only difference is the MULTILEX position which splits the two types of Conjunctions into two different categories, *coordinators* and *subordinators*: however, no mapping problems arise. The same choice was made by the TEI. [Monachini e Calzolari, 1996]

Na aplicación ás distintas linguas, os atributos opcionais xenéricos propostos só se empregan en inglés, alemán, danés ou sueco mentres que non se aplican en italiano, catalán, francés ou portugués; en castelán non se usan no nivel morfolóxico senón na identificación de constituíntes no nivel de análise sintáctica.

1.2.3.9. Numeral

En Leech e Wilson (1996) admítase que os numerais, polo seu comportamento, poden integrarse no interior doutras clases de palabras; así, no modelo portugués, os cardinais inclúense nas categorías Pronome/Determinante e os ordinais na clase Adxectivo. Con todo, deixan clara a súa postura ao afirmar que na categoría Numeral é posible especificar a función dos distintos elementos cun atributo.

En Monachini e Calzolari (1996) recoñécese o tratamento diverxente que presentan os numerais nos sistemas analizados pero, pese a permitir incluílos como subtipos das clases Pronome, Determinante e Adxectivo, a proposta para a clase Numeral non difire da realizada con anterioridade en Leech e Wilson (1996), que é como segue:

L0	Numeral
L1	Tipo: cardinal, ordinal
	Xénero: masculino, feminino, neutro
	Número: singular, plural
L2	Función: pronome, determinante, adxectivo

En Monachini e Calzolari (1996) a aplicación ao castelán non aparece recollida baixo a clase Numeral e tampouco a atopamos como subtipo do Adxectivo, Pronome ou Determinante. O portugués considera os ordinais adxectivos, mentres que os cardinais determinantes remíteos á clase Determinante e os pronominais van á de Adxectivo.

1.2.3.10. Única

A proposta de EAGLES recomenda incorporar os mínimos elementos posibles a esta clase, a cal, ademais, só debe conter fenómenos lingüísticos específicos da lingua en cuestión. Estamos, polo tanto, ante a única categoría para a que non existe ningún atributo recomendado no nivel L1; nela pásase do nivel obrigatorio (delimitación da categoría: L0) ao nivel opcional específico (trazos morfosintácticos específicos de non máis de tres linguas: L2b).

Os valores que propoñen caracterizar nesta clase son os seguintes (Leech e Wilson, 1996):

Unique-type:	1. Infinitive marker	[German <i>zu</i> , Danish <i>at</i> , Dutch, English]
	2. Negative particle	[English <i>not</i> , <i>n't</i>]
	3. Existential marker	[English <i>there</i> , Danish <i>der</i>]
	4. Second negative particle	[French <i>pas</i>]
	5. Anticipatory <i>er</i>	[Dutch]
	6. Mediopassive voice marker <i>se</i>	[Portuguese]
	7. Preverbal particle	[Greek]

As aplicacións castelá e portuguesa non están recollidas en Monachini e Calzolari (1996) e en Leech e Wilson (1996), como se observa na táboa, achégase para o portugués a posibilidade de identificar *se* como marcador de voz, opción que tamén sería válida para o castelán malia que non o fagan constar.

1.2.3.11. Residual

Igual que sucedía coa clase Única, dende EAGLES recoméndase remitir á categoría Residual os menos elementos posibles e, sempre que sexa factible, aconséllase incorporalos a algunha das outras categorías. Por exemplo, os caracteres alfabéticos varían no número, comportándose deste xeito como os substantivos comúns, motivo polo que poderían remitirse á clase Substantivo.

Reproducimos co mesmo formato a proposta de EAGLES para a clase Residual realizada en Leech e Wilson (1996) e a defendida en Monachini e Calzolari (1996) para que, a simple vista, se poidan apreciar as diferenzas:

(Leech e Wilson, 1996)			
EAG-L0	RESIDUAL		
EAG-L1	foreign word	m	sg
	formula	f	pl
	symbol	n	
	acronym		
	abbreviation		
	unclassified		

(Monachini e Calzolari, 1996)			
EAG-L0	RESIDUAL		
EAG-L1	foreign words	m	sg
	alphabetic symbols	f	pl
	formulae	n	
	acronyms		
	abbreviations		
	unclassified		
	affixes		
	shortcuts		
	fused forms		
	trunc. forms		
	comp. forms		
EAG-L2b		c	n

Segundo Leech e Wilson (1996), o tipo Inclasificado aplícase a segmentos textuais que non encaixan nos valores anteriores como, por exemplo, palabras incompletas, retrousos como *er*, *erm* nas transcricións de texto oral ou representacións escritas de onomatopeas como *dum-de-dum*.

Monachini e Calzolari (1996) non ofrecen datos sobre os diferentes valores establecidos nin recollen a aplicación para o castelán nin para o portugués; entre as linguas máis próximas á nosa figura o catalán, que acolle nesta clase os acrónimos (*SIG*) e as unidades caracterizadas con *non codificado* (NC), e o italiano que distingue aquí *estranxeirismos, abreviaturas e siglas*.

1.3. O etiquetario de XIADA

1.3.1. Clases de palabras

As categorías que delimitamos no sistema de etiquetaxe de XIADA para o galego son:

1	S [substantivo]	2	V [verbo]	3	A [adxectivo]
4	D [artigo]	5	E [demostrativo]	6	M [posesivo]
7	N [numeral]	8	G [interrogativo- exclamativo]	9	T [relativo]
10	I [indefinido]	11	R [pronome]	12	W [adverbio]
13	C [conxunción]	14	L [locución]	15	P [preposición]
16	Y [interxección]	17	Q [sinal de puntuación]	18	Z [periférica]

É evidente que o número de categorías establecidas por EAGLES (13) e por nós (18) non coincide, así como tampouco a tipoloxía de categorías na súa totalidade. A diverxencia orixínase, sobre todo, na distinta consideración de dúas das clases de palabras propostas por EAGLES: as clases P [pronome] e D [determinante].

Os motivos que nos levaron a non seguir fielmente a súa proposta esixen unha explicación.

En primeiro lugar, no referente ás categorías Pronome e Determinante, obsérvase que na proposta do estándar europeo empregan criterios morfolóxicos para a delimitación das principais clases de palabras (substantivos, verbos, adxectivos, etc.) e sintácticos para a categorización de demostrativos, indefinidos, numerais, posesivos, relativos e interrogativos. Consideran que nestes últimos é máis importante o seu funcionamento como pronomes, determinantes ou adxectivos que a clase de palabra á que tradicionalmente se remiten, información esta que secundariamente incorporan na etiqueta cun atributo que sinala o tipo de especificador, adxectivo ou pronome.

En XIADA respectamos o criterio morfolóxico, e só secundariamente coa caracterización dun atributo “Valor” clasificamos nalgunhas clases de palabras o funcionamento como nucleares ou adxacentes dos seus elementos constituíntes. Para esta decisión apoiámonos, sobre todo, en Álvarez, Regueira e Monteagudo (1986: 157-293) quen, a pesar de recoñecer que “*dentro dos pronomes distínguese entre pronomes persoais, demostrativos, posesivos, indefinidos, numerais, relativos e interrogativo-exclamativos*”, á hora de tratar cada un destes grandes grupos faino en capítulos independentes e como clases morfolóxicas distintas.

Na nosa decisión pesaron tamén outros factores importantes:

1) Exportabilidade. Se fose necesario, poderíamos converter facilmente as categorías Demostrativo, Posesivo, Indefinido, Numeral, Relativo e Interrogativo-exclamativo en subtipos da clase Pronome e Adxectivo. Como veremos máis adiante, todos os elementos integrados nalgunha destas clases inclúen información sintáctica sobre o tipo de función que poden realizar, e así caracterízanse mediante o atributo “Valor” como formas *determinantes* ou *non determinantes*. Para proceder á conversión no etiquetario destas categorías, entre outras accións, teríamos que inverter a orde dos dous primeiros atributos da cadea de etiquetaxe en cada unha destas clases, excepto no Numeral onde sería o primeiro e o terceiro (a segunda posición está reservada para indicar o tipo de numeral).

Para que se perciba con máis claridade esta hipotética conversión, imos exemplificar coa clase Demostrativo. Así, os neutros *isto, iso* e *aquilo* e as súas variantes non normativas na actualidade, *esto, eso* e *aquelo* etiquétanse como Enns (*Demostrativo, non determinante, neutro, singular*). Trataríase, daquela, de extraer do paradigma dos demostrativos os elementos marcados como Enns e categorizalos na clase Pronome co subtipo *demostrativo*: Rd00ns (*Pronome, demostrativo, caso non aplica, persoa non aplica, neutro, singular*). Isto provocaría, por unha banda, a ampliación dos valores do atributo *tipo* para darlle cabida, neste caso, aos demostrativos e, por outra, a non pertinencia dos atributos *caso* e *persoa* para todos os integrantes do paradigma do Pronome.

Para os demostrativos caracterizados como *determinante/non determinante* (o resto do paradigma, fóra o neutro) cuxa etiqueta comezaría por Ea (*Demostrativo,*

determinante/non determinante) teríamos que reduplicar a súa análise e categorizalos respectivamente na clase Adxectivo, ou crear a de Determinante *ex profeso*, e na de Pronome. Deste xeito, o elemento *este* sería caracterizado respectivamente como Rd00ms (*Pronome, demostrativo, caso non aplica, persoa non aplica, masculino, singular*) e como Ad0ms (*Adxectivo, demostrativo, grao non aplica, masculino, singular*).

Esta posible homoxeneización con outros etiquetarios provocaría, en primeiro lugar, un incremento do número de etiquetas posibles debido ás caracterizacións categoriais duplas dos determinativos que poden funcionar como núcleos e como adxacentes; en segundo lugar, a reorganización dos atributos descritivos existentes (a *persoa* ou o *caso* non son pertinentes para o Demostrativo, Indefinido, Numeral, etc.) e a adición doutros non pertinentes para os elementos orixinais pero imprescindibles para clasificar morfoloxicamente os novos membros das clases Pronome e Adxectivo (o *número do posuidor*, por exemplo, só é pertinente para os posesivos, sexa o seu valor determinante ou non).

2) Congruencia nos paradigmas. Agás para os numerais, cada unha das clases implicadas está constituída por un inventario pechado. Coa clasificación que propón EAGLES, elementos pertencentes ao mesmo paradigma remitiríanse a categorías distintas. É o que sucede, por exemplo, no etiquetario do Institut Universitari de Lingüística Aplicada (IULA) da Universitat Pompeu Fabra de Barcelona para o catalán e o castelán (Morel et al., 1998: 26, 30, 52 e 56):

meu	Especificador, possessiu, un poseïdor, primera persona, masculí, singular
mon	Adjectiu, possessiu, un poseïdor, primera persona, masculí, singular
mío	Especificador, posesivo, un poseedor, primera persona, masculino, singular
mi	Adjetivo, posesivo, un poseedor, primera persona, género pendiente, singular

mentres que no lexicón de XIADA se remiten á mesma categoría, e só secundariamente se dá conta do seu comportamento sintáctico:

meu	Posesivo, determinante/non determinante, 1ª singular, masculino, singular
mi	Posesivo, determinante, 1ª singular, masculino/feminino, singular

3) Coherencia nos atributos. O establecemento das categorías Determinante e Pronome, con criterios sintácticos, provoca a non pertinencia dun ou varios atributos específicos para cada unha das clases. Así, por exemplo, no etiquetario do IULA atopamos o caso máis rechamante na categoría Pronome, onde dos seis atributos aos que debe responder cada membro desta clase, só un é pertinente para definir Pronomes indefinidos, demostrativos e cuantificadores (Morel et al., 1998: 35-43):

REEN3MP	Pronom, personal, feble, nominatiu, terceira persona, masculí, plural (<i>els</i>)
REO-3MS	Pronom, personal, fort, terceira persona, masculí, singular (<i>ell</i>)
RD-----	Pronom, demostratiu (<i>açó</i>)
RF-----	Pronom, cuantificador (<i>res</i>)
RN-----	Pronom, indefinit (<i>algú</i>)

en troques, en XIADA, ao considerar cada clase por separado, todos os pronomes persoais responden aos mesmos atributos e o mesmo sucede cos elementos das clases Demostrativo e Indefinido para as cales só son pertinentes os atributos *valor*, *xénero* e *número*:

[Rtr3ms]	Pronome, tónico, nominativo/preposicional, 3ª, masculino, singular (<i>el</i>)
[Raa3ms]	Pronome, átono, acusativo, 3ª, masculino, singular (<i>o</i>)
[Enns]	Demostrativo, non determinante, neutro, singular (<i>isto</i>)
[Inas]	Indefinido, non determinante, masculino/feminino, singular (<i>cadaquén</i>)

4) Exhaustividade na análise. Como se pode deducir do exposto ata agora, o delimitar máis clases ca as propostas por EAGLES non impide que poidamos individualizar a través da etiqueta moitos dos elementos pertencentes ao mesmo paradigma, tanto no que respecta á súa flexión de xénero e número como ao seu funcionamento como formas nucleares e/ou modificadoras. Non todos os etiquetarios chegan a este nivel de análise. Por exemplo, o empregado para etiquetar o *Corpus Anotado do Português Europeu*²⁹ proporciona información sobre a categoría morfosintáctica principal, pero non dá conta, en cambio, de atributos como xénero, número ou grao, e tampouco distingue, nas clases que nos ocupan, entre Pronome e Determinante, senón que emprega unha clasificación semellante á nosa para “evitar uma

²⁹ Manual de anotación dispoñible en http://www.clul.ul.pt/sectores/linguistica-de-corpus/manual_annotacao.pdf [consultado o 12/03/2010]

duplicação das categorias e também para evitar níveis de erro elevados na anotação automática” (p. 4).

Acabamos de ver os motivos que nos levaron a non respectar as clases morfolóxicas de EAGLES Determinante e Pronome. Pero, a maiores, hai outras dúas categorías nas que diferimos.

Unha delas é a clase “U” [única/inclasificable]; cremos que non ten reflexo na gramática do galego a existencia dunha categoría tal polo simple feito de que exista un número moi reducido de elementos integrantes desa clase de palabra. Ademais, tampouco vemos conveniente crear unha nova categoría para dar cabida a aquelas palabras cuxa clasificación non aparece recollida nas gramáticas e/ou dicionarios e nos crea abondosas dúbidas. É preferible aproveitar a existencia da clase residual, engadir un valor novo e integrar aí os elementos que non sabemos como catalogar.

Por último, clasificamos as conxuncións, preposicións e adverbios que están constituídos por unidades multipalabra nunha clase á que denominamos “Locución”. Dende EAGLES, Monachini e Calzolari recomentan integrar as unidades multipalabra nas correspondentes categorías gramaticais simples, pero deixan aberta a opción de categorizalas doutro modo:

In general, EAGLES recommends handling multiwords as belonging to a single grammatical category, and contractions as two separate grammatical categories, but the option of a different treatment is left open. [Monachini e Calzolari, 1996]

Non habería ningún problema en recategorizar os elementos pertencentes a estas dúas clases e homoxeneizalas coa proposta de EAGLES posto que a información está contida no etiquetario, só que a un nivel distinto.

En resumo, tal como recomenda EAGLES, o sistema de etiquetación de XIADA presenta unha estrutura xerárquica na que no primeiro nivel se identifica a categoría morfolóxica e nun segundo os atributos gramaticais pertinentes que constitúen o conxunto básico de atributos para cada categoría.

Esta división en dous niveis percíbese con claridade no formato da cadea de etiquetaxe: a letra que identifica a clase de palabra á que pertence o elemento vai en maiúscula e o valor dos atributos represéntase con minúsculas. Ex.: Scms (*Substantivo, común, masculino, singular*).

A distribución dos atributos dentro de cada categoría é simétrica, de tal xeito que todos os elementos pertencentes a unha categoría dada posúen o mesmo número de atributos e os seus valores ocupan a mesma posición dentro da cadea resultante. Así, os pronomes illados *nós* e *nos* caracterízanse no lexicón como:

nós: [Rtr1ap] *Pronome, tónico, nominativo/preposicional, 1ª, masculino/feminino, plural, (lema) nós*

nos: [Raa1ap] *Pronome, átono, acusativo, 1ª, masculino/feminino, plural, (lema) nos*

nos: [Rad1ap] *Pronome, átono, dativo, 1ª, masculino/feminino, plural, (lema) nos*

nos: [Raf1ap] *Pronome, átono, formante léxico, 1ª, masculino/feminino, plural, (lema) nos*

Hai, así e todo, etiquetas nas que en consonancia co grao de especificación ao que queiramos chegar ou ben porque a atribución dun valor inequívoco sexa imposible podemos neutralizar o valor dalgún ou de varios atributos. Marcamos esta ambigüidade co valor “a” que representará o xénero, número, persoa, posuidor ou valor dependendo da clase de palabra e da posición que ocupe na cadea de etiquetaxe. Os seguintes exemplos mostran a ambigüidade no *número*, *valor* e *número do posuidor* e mais *persoa* nas categorías Substantivo, Posesivo e Verbo, respectivamente:

atlas: [Scma] Substantivo, común, masculino, singular/plural, (lema) atlas

seu: [Ma3ams] Posesivo, determinante/non determinante, 3ª, singular/plural, masculino, singular, (lema) seu

cantaría: [Vcia0s] Verbo, pospretérito, indicativo, 1ª/3ª, xénero non aplica, singular, (lema) cantar

Por outra banda, non todos os elementos pertencentes a unha mesma categoría morfolóxica responden positivamente aos atributos que a caracterizan. Así, na categoría verbal, o atributo xénero só é un trazo pertinente para a caracterización do participio. Para o resto de elementos do paradigma verbal a aplicación do atributo xénero non é pertinente. Por este motivo creamos un valor “0” co significado de “Non aplicable”. O seguinte trío de formas verbais analizadas ilustra este valor:

traiamos: [Vps10p] Verbo, presente, subxuntivo, 1ª, xénero non aplica, plural, (lema) traer

traendo: [V0x000] Verbo, tempo non aplica, xerundio, persoa non aplica, xénero non aplica, número non aplica, (lema) traer

traído: [V0p0ms] Verbo, tempo non aplica, participio, persoa non aplica, masculino, singular,
(lema) traer

Evidentemente, de non agrupar todos os valores posibles dun elemento calquera nunha única etiqueta, o analizador tería que dar todas as posibles etiquetas e o proceso informático sería moi custoso. Con todo, a aglutinación de definicións impediría desambiguar en numerosos casos polo que, na elaboración da etiquetaxe, se impuxo unha combinación dos dous criterios. Por unha banda, foi preciso manter a caracterización morfolóxica co atributo “a” ben porque morfoloxicamente se trate da análise correcta ben porque non sexa posible decidirse por unha ou outra opción. Por outra banda, non importa que o proceso sexa máis custoso, cómpre desagregar as análises para poder desambiguar.

1.3.2. Atributos

1.3.2.1. Substantivo (S)

Segundo a *Gramática galega* (Álvarez, Regueira e Monteagudo, 1986: 53) “*os sustantivos son palabras que denotan entidades, sexan estas seres vivos, cousas inanimadas ou conceptos abstractos. Teñen marcas de seu e non condicionadas polas necesidades de concordancia dentro do enunciado. A súa función básica é a de seren núcleo da frase nominal*”.

No etiquetario de XIADA os atributos que caracterizan os substantivos son os recomendados por EAGLES: **tipo**, **xénero** e **número**.

Así, calquera substantivo debe responder a estes tres atributos:

tipo	xénero	número
c (común)	m (masculino)	s (singular)
p (propio)	f (feminino)	p (plural)
	a (masculino/feminino)	a (singular/plural)
	0 (non aplica)	0 (non aplica)

Exemplos:

Teño casa, coche e paraugas [Scfs], [Scms], [Scma]

Pedro e María viven en Lugo [Spm0], [Spf0], [Sp00]

Gústalles a Cocacola [Spfs]

O valor “a” representa o sincretismo formal que presentan algúns substantivos no xénero, entre o masculino e o feminino (*o atleta/a atleta*), e no número, entre o singular e o plural (*un paraugas/varios paraugas*) e correspóndese cos valores que EAGLES denomina, respectivamente, “común” e “invariable”:

The values *common* (*c*) and *invariant* (*n*), presented in the Italian and Spanish tables, for the attributes **Number** and **Gender** respectively, raise an important issue: they are not a gender or a number, but they are a signal that gender or number cannot be determined from the form of the word alone. [Monachini e Calzolari, 1996]

Tal como se admite en EAGLES, “c” e “n”, ou no noso caso “a”, non son un xénero ou un número, pero si posibilitan marcar a indeterminación de xénero ou número tanto en formas illadas como naqueles casos nos que o contexto non nos permita determinalo.

Por razóns mnemotécnicas, no noso etiquetario a ambigüidade é caracterizada sempre co mesmo carácter: “a”. Así, como iremos vendo na descrición dos distintos atributos, o valor “a” representará os atributos de xénero, número, persoa, posuidor ou valor dependendo da clase de palabra e da posición que ocupe na cadea de etiquetaxe.

Como se pode observar na táboa, engadímoslles aos atributos *xénero* e *número* un valor non recoñecido para o Substantivo en EAGLES; trátase do valor “0” e con el indicamos que o atributo ao que se refire non aplica nesa unidade léxica concreta. Prevedemos a súa utilización na codificación dos nomes propios. Nos topónimos, por exemplo, consideramos que non é pertinente nin o xénero nin o número porque dependendo do xenérico no que pensemos (poboación, cidade, pobo, lugar, etc.) a concordancia variará:

En titular interior, a cinco columnas, o titular redúcese ata afirmar que "Vigo é máis caro que Marbella". [CORGA: *ANT2002_10_24/10*]

Ourense é máis urbana. Lugo está máis isolada, depende máis do seu carácter de cidade cabeceira dunha grande comarca rural. [CORGA: Suso de Toro, *Camilo Nogueira e outras voces*, Xerais, 1991]

Nos antropónimos distinguimos entre femininos (*María, Teresa Vázquez*) e masculinos (*Pedro, Manuel Fernández*), pero neles, por regra xeral, non aplicamos o

número a non ser “*cando pasan a usarse á maneira de adxectivos, ben para referirse a toda unha familia (os Andrades), ben usados metonimicamente (Seica ten tres Laxeiros) ou ben como calificativos atribuíndolle a un individuo as cualidades que se recoñecen á persoa portadora do nome propio (Estamos un pouco faltos de Castelaos e de Bóvedas)*” (Álvarez, Regueira e Monteagudo, 1986: 66).

Como acabamos de comprobar, non realizamos ningún tipo de subespecificación semántica nin nos substantivos comúns nin nos propios. É certo que, para os comúns, EAGLES propoñía distinguir entre *contables* e *non contables* pero, se queremos ser exhaustivos na análise, por que non tamén entre *concretos* e *abstractos* ou *individuais* e *colectivos*? E o mesmo se podería dicir dos propios, por que non distinguir entre eles *antropónimos*, *topónimos*, *organismos*, *empresas*, etc. A razón pola que non levamos adiante unha descrición tan detallada é dobre; por unha banda, non entraba no noso obxectivo a consideración de criterios semánticos e, pola outra, concluímos que o proveito que se lle podía sacar a unha caracterización semellante non compensaba o tempo e recursos que había que empregar para levala a cabo.

1.3.2.2. Verbo (V)

Segundo Álvarez, Regueira e Monteagudo (1986: 295) “*o verbo caracterízase fundamentalmente fronte ás outras clases de palabras pola súa capacidade para expresa-lo tempo e o modo da cláusula, ademais do número e persoa do suxeito. A súa función obrigatoria é a de núcleo da Frase Verbal*”.

Igual que no substantivo, tampouco aquí imos atender ás subespecificacións que se poidan derivar de criterios sintácticos ou semánticos e así non teremos en conta se o verbo é *transitivo* ou *intransitivo*; *reflexivo*, *recíproco* ou *impersoal* ou, finalmente, *incoativo*, *frecuentativo*, *perfectivo* ou *imperfectivo*.

Os atributos que caracterizan a clase de palabra *Verbo* no noso sistema son os seguintes:

tempo	modo	persoa	xénero	número
p (presente)	i (indicativo)	1 (primeira)	m (masculino)	s (singular)
i (copretérito)	s (subjuntivo)	2 (segunda)	f (feminino)	p (plural)
l (antepretérito)	m (imperativo)	3 (terceira)	0 (non aplica)	0 (non aplica)
f (futuro)	f (infinitivo)	a (primeira e terceira)		
e (pretérito)	x (xerundio)	0 (non aplica)		
c (pospretérito)	p (participio)			
0 (non aplica)				

Adoptamos a terminoloxía empregada nas *Normas ortográficas e morfolóxicas do idioma galego* (RAG/ILG, 1995: 105-107), en diante *Normas*, para denominar os distintos tempos verbais: *presente* (canto, cante), *copretérito* (cantaba), *antepretérito* (cantara), *futuro* (cantarei, cantar), *pretérito* (cantei, cantara, cantase) e *pospretérito* (cantaría).

Incluimos as formas en *-ra* no pretérito de Subxuntivo porque no galego actual, tanto no falado coma no escrito, sexa este último literario ou non³⁰, é usual a súa presenza, malia que na actualidade as *Normas* (RAG/ILG, 2003: 111) recomendan “*manter ben discriminados os usos dos dous tempos (-ra coma antepretérito de indicativo e -se coma pretérito de subjuntivo)*”.

Tradicionalmente, nas gramáticas de linguas próximas á nosa acostúmase facer unha distinción xenérica entre *formas finitas* ou *persoais* e *formas infinitas* ou *non persoais*. Discrimínase así con este trazo o infinitivo, xerundio e participio fronte ao resto de formas do paradigma verbal. No deseño do etiquetario decidimos non telo en conta por dous motivos; primeiro, en galego existe un tempo denominado infinitivo conxugado, con plena vixencia na lingua actual, que, como o seu nome indica, presenta desinencias de número e persoa; por se non bastase con isto, tamén se constata, aínda que moito máis esporadicamente, a presenza do xerundio conxugado³¹, con formas diferenciadas para a primeira e segunda persoas de plural. A existencia do infinitivo e do xerundio conxugados invalida a distinción tradicional entre *formas persoais* e *formas*

³⁰ En Álvarez, Regueira e Monteagudo (1986: 381), recóllese este uso das formas en *-ra* no galego falado e na lingua literaria. Nas *Normas* de 1995 (RAG/ILG, 1995: 111) recoñécese o funcionamento das formas en *-ra* como pretérito de subjuntivo, apuntando que incluso son as únicas posibles nalgúns construcións, sen que se engada ningunha particularidade máis sobre o seu uso.

³¹ A primeira persoa do plural do xerundio conxugado concorre en 16 ocasións na versión 1.4 do CORGA, mentres que non se rexistra ningún caso da segunda de plural [consultado o 25/05/2010].

non persoais. En segundo lugar, a oposición *forma persoal* fronte a *forma non persoal* súplese cos atributos de persoa e número; é dicir, se quixésemos manter o trazo *persoal* vs. *non persoal* estaríamos reduplicando información en canto que toda forma verbal debe caracterizarse con respecto á persoa e ao número.

No noso etiquetario agrupamos infinitivo conxugado e infinitivo nun único infinitivo, sabendo que a diferenza saltará no atributo persoa. Procedemos de igual xeito co xerundio e xerundio conxugado. Ex.: *Antes de saírmos temos que preguntar se nos deixan*:

saírmos: [V0f10p] Verbo, tempo non aplica, infinitivo, primeira, xénero non aplica, plural, (lema) saír

preguntar: [V0f000] Verbo, tempo non aplica, infinitivo, persoa non aplica, xénero non aplica, número non aplica, (lema) preguntar

Seguimos, en consecuencia, a proposta de EAGLES e incluímos infinitivo, xerundio e participio como valores modais malia saber que non son modos, porque así evitamos ter que empregar outro trazo para caracterizar estas formas verbais e non complicamos a análise excesivamente. En todo caso, con este sistema de etiquetaxe podemos individualizalos cando se queira e homoxeneizar a súa etiquetación coa doutro sistema calquera.

Tradicionalmente o imperativo considérase un modo e así o mantemos. Non cremos oportuno, en cambio, estimar para o denominado *condicional* un valor modal. Non é este o lugar para estendernos sobre as distintas teorías lingüísticas acerca de cantos modos e cales se recoñecen nas tradicións hispánica ou portuguesa. Baste sinalar que seguimos o único modelo conxugacional prescriptivo existente na época de elaboración do etiquetario de XIADA: as *Normas ortográficas e morfolóxicas do idioma galego* (RAG/ILG, 1995).

No nivel morfolóxico da aplicación ao portugués de GENELEX (Monachini e Calzolari, 1996) diferencian no atributo *persoa* entre a *deíctica* e a que impón a concordancia; así para cada tempo persoal distinguen dúas formas máis, as cortesés de singular e plural. Vémosto aplicado, por exemplo, no presente de indicativo:

01	Ind-Pr-1-1-S	como	02	Ind-Pr-2-2-S	comes
03	Ind-Pr-2-3-S	come	04	Ind-Pr-3-3-S	come
05	Ind-Pr-1-1-P	comemos	06	Ind-Pr-2-2-P	comeis
07	Ind-Pr-2-3-P	comem	08	Ind-Pr-3-3-P	comem

Pola nosa banda estamos de acordo en que os pronomes *vostede*, *vostedes* se refiren sempre ao interlocutor, unha segunda persoa polo tanto, pero a efectos gramaticais levan o verbo en terceira persoa, singular ou plural, e funcionan como un pronome de terceira, razón pola que lles corresponden as formas átonas *o(s)*, *a(s)* para o complemento directo, *lle(s)* para o indirecto, *si* para o pronome reflexivo tónico e *seu* para o posesivo:

Pídolles a vostedes que, por favor, non lles poñan ningún impedimento para o traballo e para seguiren adiante, porque eles non son culpables; [CORGA: Santiago Álvarez, *Memoria de Guerrilla*, Xerais, 1991]

Despois de todo eu a vostedes non os coñezo de nada. [CORGA: Manuel Forcadela, *Fóra de xogo*, Edicións do Cumio, 1993]

Queda moito por facer", dixo o ministro, "pero confío en vostedes, señores, no seu sacrificio e no seu patriotismo". [CORGA: Víctor F. Freixanes, *A cidade dos Césares*, Xerais, 1993]

¡Volve vostede en si! [CORGA: Xavier Manteiga, *Morrer na herba*, Edicións Positivas, 1995]

Álvarez, Regueira e Monteagudo (1986: 161-162) inclúen as formas tónicas para a cortesía (*vostede*, *vostedes*) nas pertencentes á segunda persoa, aínda que advirten das concordancias coas terceiras persoas, tanto verbais como pronominais. Alarcos (1994: 76) recólleas como substitutas cortesés dos pronomes persoais de segunda persoa, malia identificar o seu comportamento gramatical coa terceira persoa. As *Normas* (RAG/ILG, 1995: 91) só indican que “as formas de cortesía son *vostede* e *vostedes*, que existen verbo en terceira persoa de singular ou plural, respectivamente”. Cunha e Cintra (1984: 278, 292-294) denomínanos “pronomes de tratamento” e coinciden na súa definición coa da gramática galega, mentres que o *Esbozo de una nueva Gramática de la lengua española* (RAE, 1986: 343) recólleas no paradigma de tratamento da 3ª persoa.

En xeral, o que se detecta nas consideracións anteriores é a confusión entre a persoa gramatical (3ª) coa sinalada pola realidade extralingüística ou deíctica (2ª). Os morfemas de número-persoa das formas verbais empregadas cos pronomes de cortesía son os correspondentes á 3ª, non hai uns específicos para a 2ª cortés:

Número	Persoa	Morfema xeral	Alomorfos
Singular	1 ^a	Ø	-o (canto) -i (cantei, cantarei) -n (collín)
	2 ^a	-s	-ches (cantaches) Ø (canta)
	3 ^a	Ø	-u (cantou)
Plural	1 ^a	-mos	-mo (sentámonos)
	2 ^a	-des	-stes (cantastes) -de (cantade)
	3 ^a	-n	-ron (cantaron)

A non consideración na conxugación verbal de formas corteses implica que o número de constituíntes por paradigma varía considerablemente do noso etiquetario (70 formas verbais) ao do portugués empregado en GENELEX (92 formas) ou ao do empregado polo Seminario de Lingüística Informática (SLI) da Universidade de Vigo co corpus TECTRA (Aguirre Moreno et al., 2002b: 212-215) (92 formas).

A maiores dos atributos *tempo*, *modo*, *persoa*, *xénero* e *número*, tal como recomenda a proposta de EAGLES³², pensamos na posibilidade de dar conta tamén dos pronomes enclíticos que se amalgaman coa forma verbal, só que neste novo atributo temos que dar cabida a unha característica peculiar do galego: a presenza de segundas formas do artigo con verbos e pronomes átonos.

No galego, como todos sabemos, adxunta a unha forma verbal rematada en *-r* ou *-s* pode ir o alomorfo do artigo *lo(s)*, *la(s)*, separándose os dous elementos cun guión. Este alomorfo tamén pode acompañar os pronomes enclíticos *nos*, *vos* e *lles*.

Fronte a linguas coma o castelán na que actualmente empregan a énclice co imperativo, infinitivo e xerundio e con non máis de dous clíticos, o galego móstrase máis complexo. Por unha banda, a única forma verbal á que non pode ir nunca enclítico un pronome é o participio (Álvarez, Regueira e Monteagudo, 1986: 203). Pola outra, en galego, ademais de ser habitual a abundancia de secuencias de clíticos, pois concorren non só *se* e as formas de acusativo e dativo que funcionan como complemento directo e indirecto, senón tamén formas de dativo que cumpren a función sintáctica de dativo de interese ou de solidariedade, poden darse ata catro dativos compatibles coa presenza de *se* (Álvarez, 1994: 247-265).

³² *Clitic*. This Boolean-valued feature is introduced at language-specific level to encode the presence in the verbal form of the clitic. [Monachini e Calzolari, 1996]

Ata o ano 2005, na listaxe de etiquetas, a información referente á presenza de pronomes átonos ou artigo amalgamados ao verbo aparecía recollida na última posición da cadea de etiquetaxe da forma verbal cun díxito que indicaba o número de pronomes e/ou artigo en énclice:

0 = ningún pronome nin artigo

1 = 1 pronome ou artigo

2 = 2 pronomes ou pronome e artigo

3 = 3 pronomes ou 2 pronomes e artigo

4 = 4 pronomes ou 3 pronomes e artigo

Así, por exemplo, recoñeciamos teoricamente para a 2ª persoa plural do presente de Indicativo (*cantades*) as seguintes posibilidades de análise:

[Vpi20p0] "Vbo. pte. ind. 2ª xen.NA pl. 0pron./art."

[Vpi20p1] "Vbo. pte. ind. 2ª xen.NA pl. 1pron./art."

[Vpi20p2] "Vbo. pte. ind. 2ª xen.NA pl. 2pron./art."

[Vpi20p3] "Vbo. pte. ind. 2ª xen.NA pl. 3pron./art."

[Vpi20p4] "Vbo. pte. ind. 2ª xen.NA pl. 4pron./art."

Ao incluír esta información, estabamos multiplicando por cinco a análise de cada unha das formas verbais que poden levar posposto un ou varios pronomes e/ou artigo e, en consecuencia, elevabamos o número de etiquetas posibles a unhas 700, aproximadamente.

No ano 2005-2006, cando acometemos a adaptación do lexicón á nova normativa do galego, decidimos descartar este atributo, posto que non achegaba ningunha información morfosintáctica e a súa rendibilidade era escasa. Dende entón, unha forma verbal con pronomes enclíticos ou cunha segunda forma de artigo, como veremos máis adiante, trátase do mesmo xeito ca unha contracción.

Exemplos de anotación verbal:*Gústannles as novelas*gústannles³³:gustan [Vpi30p gustar] = Verbo, presente, indicativo, 3ª, xénero non aplica, plural, (lema) gustar.lles [Rad3ap lle] = Pronome, átono, dativo, 3ª, masculino/feminino, plural, (lema) lle.*Preferiría que non viñeses*preferiría: [Vcia0s preferir] = Verbo, pospretérito, indicativo, 1ª/3ª, xénero non aplica, singular, (lema) preferir.viñeses: [Ves20s vir] = Verbo, pretérito, subxuntivo, 2ª, xénero non aplica, singular, (lema) vir.*Ninguén leva aturado tanto coma min*leva: [Vpi30s levar] = Verbo, presente, indicativo, 3ª, xénero non aplica, singular, (lema) levar.aturado: [V0p0ms aturar] = Verbo, tempo non aplica, participio, persoa non aplica, masculino, singular, (lema) aturar.**1.3.2.3. Adxectivo (A)**

Os adxectivos son palabras que expresan cualidades atribuídas a obxectos denotados polos sustantivos. Formalmente non se distinguen dos sustantivos e coma eles admiten morfemas de xénero e número, mais coa diferenza de que os do adxectivo están condicionados gramaticalmente. (...) A cualidade sinalada polo adxectivo pode ser de moi diversos tipos: accidental ou permanente, implicada polo propio significado do sustantivo, e por tanto compartida por tódolos individuos da súa clase, ou particularizadora dese individuo fronte ós outros da mesma clase designados polo sustantivo. [Álvarez, Regueira e Monteagudo, 1986: 69]

Tanto a *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986: 69-70) como a *Nova Gramática para a aprendizaxe da lingua* (Costa Casas et al., 1988: 70-71)

³³ Na actualidade, as unidades gráficas que posúen máis dun elemento gramatical tipo *gústannles* descompóñense en constituíntes; neste caso, un para a análise da forma verbal e outro para a do pronome enclítico:

Análise unidade ▾
Unidade
gústannles
Constituínte ▾
Forma gustan Etiqueta Vpi30p Lema gustar
Constituínte ▾
Forma lles Etiqueta Rad3ap Lema lle

distinguen dous tipos de adxectivos: os especificativos e os explicativos ou epítetos. En ningún momento, no capítulo dedicado a esta clase, se fai referencia aos denominados determinativos; estes son considerados pronomes en ambas as dúas.

Pola contra, tal como describe Alarcos (1994), en castelán, ademais de clasificar os adxectivos en especificativos e explicativos, acostúmase tamén distinguir entre adxectivos *cualificativos* e adxectivos *determinativos*, agrupándose nestes últimos os demostrativos, posesivos, numerais, indefinidos e relativos. Como podemos constatar, a proposta de clasificación de EAGLES para esta clase está ben fundamentada na tradición gramatical. Non obstante, fronte ás reflexións de Alarcos (1994: 82-84), en Monachini e Calzolari (1996) non se ten en conta que o comportamento dos cualificativos e dos determinativos en canto á súa posición e á do grupo nominal que modifican diverxe o duns e o doutros. Así, seguindo a Alarcos, os primeiros admiten liberdade de posición entre si e respecto do núcleo substantivo que delimitan (*esta fresca e saborosa auga; esta auga fresca e saborosa; esta saborosa auga fresca*), mentres que os determinativos presentan restricións en canto á súa posición dentro do grupo e, na presenza doutro adxectivo no mesmo grupo, esixen estar antepostos e nunca poden ir inmediatamente pospostos a este (*esta saborosa auga* pero non *saborosa esta auga*).

O propio Alarcos (1994: 84) remata afirmando que a partir destes comportamentos diferentes conclúese que hai dous tipos de adxectivos “*separados funcionalmente por su diversa posibilidad de ordenación entre sí y respecto del sustantivo al que acompañan*”: os cualificativos e os determinativos. Esta conclusión, xunto coa descrición das dúas gramáticas galegas anteriores, foi determinante para desbotar a proposta de delimitación da clase Adxectivo de EAGLES e considerar cada tipo como categoría independente.

Ademais, tal como indicamos xa na xustificación xeral da delimitación de categorías realizada no noso etiquetario, o evitar introducir atributos pertinentes só para un número limitado de elementos determinativos, como son *persoa* ou *número do posuidor* na clase Adxectivo, proporcionalle consistencia e congruencia ao *tag-set* de XIADA. Na proposta de EAGLES admítese a incoherencia, pero non se lle concede maior importancia:

The fact of having indicative adjectives motivates the presence in this table of features such as **Number of Possessor** and **Person**.

Each language-specific application should specify clearly in which category the indicative adjectives are treated in order for cross-linguistic comparisons to be made possible. [Monachini e Calzolari, 1996]

No etiquetario de XIADA calquera adxectivo debe responder a estas tres características:

grao	xénero	número
c (comparativo)	m (masculino)	s (singular)
s (superlativo)	f (feminino)	p (plural)
0 (non aplica)	a (masculino/feminino)	a (singular/plural)

Exemplos:

Este amable señor é mellor persoa ca ti [A0ms], [Acfs]

Préstame a máxima atención [Asfs]

Atleta agradable [A0as]

Atleta agradabilísimo [Asms]

Introducimos como un atributo máis o *grao*, recomendado por EAGLES, que nos permitirá distinguir os comparativos e superlativos sintéticos dos analíticos. Marcará como comparativos *mellor*, *peor*, *maior*, *menor* e *meirande* e como superlativos os irregulares *óptimo*, *pésimo*, *máximo* e *mínimo* e todos os regulares sintéticos. Consideramos que o termo non marcado no *grao* é o valor positivo, polo que, por defecto, o *grao* non será aplicable a non ser que sexa para marcar un comparativo ou superlativo sintético.

Non incluimos no noso etiquetario os atributos xenéricos opcionais de EAGLES *uso* e *tipo de modificación* cuxo establecemento, constatamos, responde a criterios sintácticos. Así, distínguense segundo o *uso* adxectivos *atributivos* (cando o adxectivo modifica un substantivo no interior dunha frase substantiva) e *predicativos* (cando o adxectivo funciona como complemento dun verbo copulativo, é dicir atributo ou como predicativo). Segundo o *tipo de modificación* os adxectivos poden ser *premodificadores* e *posmodificadores*. A nivel de palabra, sen que a unidade léxica estea inmersa nunha

secuencia, os únicos adxectivos que poderíamos caracterizar como *premodificadores* en galego son aqueles que sofren apócope: *gran*, *mal* e *san*, cuxas formas plenas son respectivamente *grande*, *malo* e *santo*.

Teoricamente, no lexicón non poderíamos empregar o atributo *uso* e o valor *premodificador*, a priori, só se usaría para marcar tres formas (*gran*, *mal* e *san*), co que resultaría pouco operativo³⁴. Non aplicamos estas distincións sintácticas no corpus, onde si sería posible realizalas, porque preferimos manter un único etiquetario para o lexicón e para etiquetar o corpus. Ademais, as ocorrencias dos tres únicos casos claros de *premodificadores* son igualmente recuperables no *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq)³⁵ mediante a consulta combinada de forma (*gran*, *mal*, *san*) e lema (*grande*, *malo*, *santo*). Non obstante, teremos en conta a recomendación de EAGLES se algún día acometemos a construción dun *treebank*.

1.3.2.4. Artigo (D)

Tanto en Leech e Wilson (1996) como en Monachini e Calzolari (1996) propúgnase tratar o Artigo como categoría independente, aínda que se acepta a súa inclusión na clase dos Determinantes se o sistema de etiquetación permite individualizalo:

Articles constitute a separate category, as is found in many lexicons and annotation practices, even though, as already pointed out, in others (e.g. UPenn) they are found incorporated in the Determiner category.

The separation of the two categories is not a problem, since, as Articles form a very restricted closed class, they can also easily be distinguished from Determiners in the latter cases and no serious mapping problem arises. [Monachini e Calzolari, 1996]

En XIADA seguimos a proposta de EAGLES á hora de caracterizar o Artigo na súa consideración de categoría propia e nos atributos que o describen. Loxicamente, non adoptamos o valor *neutro* para o xénero porque en galego este non é pertinente no artigo.

³⁴ Normalmente, en galego o adxectivo segue ao substantivo, mais en moitos casos pode adoptar unha posición anterior ou posterior a este, co que non poderíamos aplicar o *tipo de modificación* sen antes estudar a fondo o adxectivo e a súa posición e, aínda así, a maioría deles presentaría ambigüidade neste atributo.

³⁵Véxase <http://corpus.cirp.es/corgaetq>

Como podemos ver na táboa descritiva, o Artigo non presenta ningunha complicación para ser caracterizado morfoloxicamente:

tipo	xénero	número
d (determinado)	m (masculino)	s (singular)
i (indeterminado)	f (feminino)	p (plural)

Exemplos:

Os nenos [Ddmp]

Había *unha* vez [Difs]

Non quere come-lo caldo [Ddms]

Non obstante, o paradigma do artigo si provocará numerosas complicacións nas fases seguintes. Primeiro, todas as formas en xeral debido á cantidade de contraccións nas que están implicadas. Logo, as denominadas segundas formas do artigo determinado (*lo, la, los, las*) van repercutir na elaboración da gramática formal que se empregue para construír o lexicón na clase verbal; e, finalmente, na desambiguación, serán as formas do artigo indeterminado (*un, unha, uns, unhas*) as que causen dificultades para separar os seus usos dos do numeral e do indefinido.

1.3.2.5. Demostrativo (E)

O demostrativo é a única categoría do galego na que o xénero neutro sobrevive. Contribúe a definir as formas normativas *isto, iso* e *aquilo* e as non normativas dende o 2003 *esto, eso* e *aquelo*.

Os atributos tidos en conta nesta categoría son:

valor	xénero	número
d (determinante)	m (masculino)	s (singular)
n (non determinante)	f (feminino)	p (plural)
a (determinante/non determinante) ³⁶	n (neutro)	

³⁶ O valor inespecificado *determinante/non determinante* nas clases Demostrativo, Posesivo, Numeral, Interrogativo-exclamativo e Relativo só se emprega na xestión interna do lexicón; na etiquetación de corpus ou no léxico xerado a partir do lexicón ese valor ambiguo desdóbrase nas etiquetas correspondentes con dous valores distintos, un como *determinante* e outro como *non determinante*. A categoría Indefinido, non obstante, como veremos no capítulo dedicado á desambiguación, mantense á marxe dos postulados anteriores.

Exemplos:

Iso paréceme unha parvada [Enns]

Esta camisa e eses pantalóns gústanme [Edfs], [Edmp]

Como podemos comprobar nos exemplos, consideramos que as formas neutras, pola súa concordancia, son singulares.

Decidimos non aplicar nesta categoría o atributo *persoa* polo mesmo motivo que xa explicamos para a non diferenciación de formas verbais corteses e non corteses: a situación dun obxecto ou dun ser no espazo ou no tempo con respecto ás tres persoas gramaticais, característica do demostrativo, é só deíctica pois, gramaticalmente, o demostrativo concorda sempre coa terceira persoa. Exs.:

Este periódico intentou onte pola tarde poñerse en contacto cos responsables destas tres empresas, sen que isto fose posible. [CORGA: CG1996-02-03/62]

A súa paga diaria por un pase alcanza os seis millóns de pesetas, pero ese é só o comezo. [CORGA: CG1996-02-23/132]

Estes cobran cifras millonarias por maltratar un animal, aqueles reciben un pagamento oneroso por procurar un pan que se chama peixe. [CORGA: ANT2005-03-03/70]

1.3.2.6. Posesivo (M)

Os posesivos caracterízanse por manter unha dobre relación coa persoa ou cousa que posúe e coa cousa ou ser posuído. Concordan en xénero e número coa cousa posuída e en número co posuidor. De aí que nesta categoría teñamos dous atributos *número*: un para o posuidor (*meu* fronte a *noso*) e outro para o posuído (*meu* fronte a *meus*).

Os atributos que caracterizan morfosintacticamente esta categoría son os seguintes:

valor	persoa	nº do posuidor	xénero	nº do obxecto posuído
d (determinante)	1 (primeira)	s (singular)	m (masculino)	s (singular)
n (non determinante)	2 (segunda)	p (plural)	f (feminino)	p (plural)
a (determinante/non determinante) ³⁶	3 (terceira)	a (singular/plural)	a (masculino/feminino) ³⁷	

³⁷ Este valor só o utilizamos no lexicón para caracterizar o posesivo de respecto; na anotación de corpus, loxicamente, levará o xénero do substantivo que determine: padre, madre, señor...

Exemplos:

É seu [Mn3ams]

Esta é a vosa oficina, aquela é a nosa [Md2pfs], [Mn1pfs]

Daquela mi padre non nos deixaba ir ás festas [Md1sms]

Merqueilles cadanseu traxe aos rapaces [Md3sms]

A caracterización do posesivo distributivo *cadanseu* no paradigma do Posesivo é, quizais, a máis problemática hoxe en día. Para a súa introdución nesta categoría fundamentámonos en Álvarez, Regueira e Monteagudo (1986: 224), onde se advirte da existencia dunhas “*formas especiais con valor posesivo-distributivo*” que se empregan “*cando se quere adxudicar a varios posuidores, considerados individualmente, un determinado obxecto para cada un, ou ben cando se lles quere distribuír un número igual ou equiparable de obxectos*”. Costa Casas et al. (1988) non o recollen e tampouco o atopamos na gramática portuguesa de Cunha e Cintra (1984). Nas *Normas* (1995: 94-95; 2003: 93-94) inclúese entre os posesivos e Ferreiro (1999) afirma no apartado centrado nos posesivos:

En galego moderno existen aínda outras formas especiais (que acompañan sempre ao obxecto posuído, nunca ao posuidor), con valor posesivo-distributivo, cadanseu(s), cadansúa(s), orixinadas pola aglutinación de cada un seu e cada un súa, respectivamente. [Ferreiro, 1999: 265]

Álvarez e Xove (2002: 482-495) clasificano no paradigma dos cuantificadores como cuantificador non comparativo de totalidade, no mesmo grupo ca *ambos(as), entrambos(as), todo(s), toda(s), cada, cada cal, cada un, cada unha, cadaquén* e no mesmo subgrupo ca *senllos(as)*. Silva (2002: 164-165) considérao un posesivo, Hermida (2004: 49-50) recólleo entre os cuantificadores, e Freixeiro (2000: 225-226; 2001: 76-78), finalmente, convéncenos coa explicación que achega sobre a súa orixe, clase morfolóxica e nivel de uso:

O posesivo xerou en galego modernamente certas construcións ou usos peculiares que, sen teren presenza na lingua histórica nen no portugués, porén non ofrecen dúbida canto á súa autoctonía. (...) Semella claro que estas formas [*cadanseu(s)* e *cadansúa(s)*] obedecen a fenómenos fonéticos propios dun rexistro popular e inclusive vulgar, característicos da fala descoñecida ou informal, que non deben tirarse do ámbito coloquial da oralidade en que se xeraron, pois entón por esa mesma vía tamén se poderían formalizar outras contraccións

populares como **pr'onha* (<para unha), **contr'onha* (<contra unha) ou **com'on* (<coma un), por exemplo, rexeitadas na norma culta a pesar de posuíren unha maior tradición literaria do que *cadanseu*, tamén xeralmente ignorada, por infrecuente ou por ser considerada como demasiado vulgar, na nosa tradición gramatical. [Freixeiro, 2001: 76-77]

Como acabamos de comprobar, a maioría de autores que o estudan consideran *cadanseu* un posesivo distributivo, razón pola que o seguiremos mantendo no paradigma da clase Posesivo. Tamén temos que mostrar o noso acordo coa afirmación realizada en Álvarez, Regueira e Monteagudo (1986: 224) sobre a maior frecuencia de uso do singular ca do plural. Consultamos a versión 1.4 do CORGA de 25 millóns de formas e estes son os resultados:

Unidade	Frecuencia
cadanseu	178
cadanseus	16
cadansúa	190
cadansúas	10
cadaseu	3
cadasua	1
cada un seu	5
cada un o seu	11
cada un súa	4
cada un a súa	6

1.3.2.7. Numeral (N)

Distinguimos na categoría Numeral co atributo *tipo* entre cardinais e ordinais. Nun primeiro momento consideramos integrar tamén os valores multiplicativo e partitivo, tal como se describe en Álvarez, Regueira e Monteagudo (1986: 265-277) e mais en Costa Casas et al. (1988: 129-136); non obstante, ao final, optamos por non tratalos como numerais senón como substantivos ou adxectivos, segundo correspondese.

Os multiplicativos, por exemplo, constitúen unha serie pechada e, cando son adxectivos, denotan que o substantivo co que concordan contén tantas unidades como aquel indica. Así, o comportamento de *dobre* en *Quería unha ración dobre* é idéntico ao de *Quería unha ración pequena* ou *Quería unha ración grande*: todos eles modifican ao substantivo co que concordan en xénero e número. Como substantivos, son sempre masculinos, van precedidos normalmente de artigo e indican cantidade por (2, 3, etc.):

Gaña o dobre/triple ca min. Só *dobre* e *triple* coas respectivas variantes que flexionan o xénero, *duplo*, *dupla*, *triplo* e *tripla*, principalmente as formas invariables, se empregan na fala cotiá, as outras pertencen a un rexistro de lingua culto (*cuádruple / cuádruplo*, *cuádrupla*; *quíntuplo*, *quíntupla*; *séxtuplo*, *séxtupla*; *décuplo*, *décupla*). Tanto é así que no subcorpus desambiguado e revisado manualmente (304.118 elementos gramaticais analizados na versión 1.0 do CORGAetq) só hai un caso de *duplo* (Scms: *Substantivo, común, masculino singular*) e 36 de *dobre*. As oracións que seguen recollen a única aparición de *duplo* e dúas de *dobre*, unha substantivo (Scms: *Substantivo, común, masculino, singular*) e outra adxectivo (A0fs: *Adxectivo, grao non aplica, feminino, singular*):

Eses 1,86 billóns son como a metade do PIB galego, ou entre o 5 e o 6 por cento dos orzamentos públicos, case o duplo dos orzamentos do Ministerio de Educación. [CORGAetq: TN1997-06/7]

Aínda que os datos establecen en 5.400 millóns o presuposto de ETA, a Ertzaintza sospeita que a cifra pode se-lo dobre. [CORGAetq: CG1996-02-15/76]

Unha vez máis son as mulleres as que viven a dobre catástrofe, a do tsunami e a deses outros efectos que se basean na discriminación que a onda non borra. [CORGAetq: ANT2005-03-03/67]

Os partitivos ou fraccionarios, pola súa banda, expresan semanticamente división dun todo en partes e designan as fraccións iguais nas que se dividiu a unidade. Unicamente *medio* e *tercio* poderían ser considerados numerais partitivos propios pois a partir de *cuarto* empréganse as mesmas formas do numeral ordinal. Nós preferimos incorporalos ás categorías Substantivo e Adxectivo, igual que fixemos cos multiplicativos, porque o seu comportamento así o indica. É máis, tanto os denominados multiplicativos como os partitivos, se os consultamos no DRAG ou no VOLGa, por exemplo, aparecen categorizados como substantivos e adxectivos, non como numerais. Un exemplo destes valores recollémolo nas seguintes oracións, onde *quinto* e *tercio* levan a etiqueta Scms (*Substantivo, común, masculino singular*) e *media* etiquétase A0fs (*Adxectivo, grao non aplica, feminino, singular*):

Bruxelas considera determinante fortalece-la estratexia comercial comunitaria, xa que a UE protagoniza un quinto do comercio mundial e un tercio dos investimentos estranxeiros directos no mundo. [CORGAetq: CG1996-02-15/95]

O representante sindical de UGT, Manuel Taboada, descoñece se este fondo para xubilacións anticipadas formará parte dun novo plan ou do xa emprendido hai tres anos, ó que se acolleron 60 empregados, anque quedan media ducia máis para este exercicio. [CORGAetq: CG1999-05-14/23]

Os denominados numerais colectivos (*par, parella, decena, ducia, quincena, vintena, etc.*) tanto en Álvarez, Regueira e Monteagudo (1986: 273) como máis tarde en Álvarez e Xove (2002: 526), malia designar un número determinado de unidades, trátanse como substantivos, análise coa que estamos de acordo e que mantemos no sistema de XIADA:

Cando fago ese milleiro de quilómetros entre A Coruña e Lisboa, observo en cambio que a partir de Tui non me sinto tan identificado. [CORGAetq: GI1996-06-30]

A lista complétase con ducias de actos de similar trascendencia que contribuíron a darlle lustre ó Xacobeo. [CORGAetq: TN1998-04/13]

Pero esa paisaxe oculta un cento de detalles nos que por forza hai que reparar porque neste momento, a máis dun mes da catástrofe, hai que buscar con máis atención, non chega coa mirada que busca as imaxes tremendas. [CORGAetq: ANT2005-03-03/66]

Os atributos que caracterizan a clase Numeral no etiquetario de XIADA son os seguintes:

tipo	valor	xénero	número
c (cardinal)	d (determinante)	m (masculino)	s (singular)
o (ordinal)	n (non determinante)	f (feminino)	p (plural)
	a (determinante/non determinante) ³⁶	a (masculino/feminino)	

Exemplos:

Quero unha sandía, dous quilos de amorodos e dúas cebolas

unha [Ncdfs] (Numeral, cardinal, determinante, feminino, singular, (lema) un)

dous [Ncdmp] (Numeral, cardinal, determinante, masculino, plural, (lema) dous)

dúas [Ncdfp] (Numeral, cardinal, determinante, feminino, plural, (lema) dous)

Son a cuarta da miña clase, logo hai tres mellores ca min

cuarta [Nonfs] (Numeral, ordinal, non determinante, feminino, singular, (lema) cuarto)

tres [Ncnap] (Numeral, cardinal, non determinante, masculino/feminino, plural, (lema) tres)

Os numerais *ordinais* caracterízanse, semanticamente, por sinalar a orde nunha serie e, gramaticalmente, por flexionar xénero e número. Así, a modo de exemplo, nas oracións que seguen, *cuarta* etiquétase en XIADA Nodfs (*Numeral, ordinal, determinante, feminino, singular, (lema) cuarto*) e *primeiras* Nodfp (*Numeral, ordinal, determinante, feminino, plural, (lema) primeiro*):

Isto coloca a Galicia na cuarta comunidade con menos investimento, por detrás de País Vasco, Navarra e Canarias, que contan cun réxime especial de financiamento. [CORGAetq: ANT2002-10-03/31]

A entidade bancaria nacional principiou a abri-las primeiras sucursais a raíz dunha aperturista lei, aprobada o 28 de xaneiro de 1856, pola que se autorizou a pluralidade de bancos de emisión. [CORGAetq: CG1996-02-09/57]

Os *cardinais* indican un número determinado de unidades e, formalmente, só *un*, *unha*, *dous*, *dúas* e mais os seus compostos e todas as centenas excepto a primeira manteñen a oposición de xénero. Nos demais *cardinais* non hai formas diferenciadas segundo o xénero; este vén determinado polo substantivo ao que se refire. Así, nos seguintes exemplos, *unha* recibe a etiqueta Ncdfs (*Numeral, cardinal, determinante, feminino, singular, (lema) un*) e *sete* Ncdfp (*Numeral, cardinal, determinante, feminino, plural, (lema) sete*):

Os reloxos atrasáronse onte unha hora e deron entrada ó horario de inverno. [CORGAetq: CG2001-10-28/4]

Sete semanas máis tarde, Brasil estoupa de novo e o espectro do cesamento de pagos ronda. [CORGAetq: ANT2002-10-31/27]

Tanto os *cardinais* como os *ordinais*, polo que respecta ao atributo *valor*, etiquétanse no lexicón co valor ambiguo *determinante/non determinante*; é na anotación de corpus onde se desfai esa anfibloxía: se determinan o substantivo que acompañan márcanse como *determinantes* e se funcionan como núcleos da frase nominal como *non determinantes*. Un caso de cada témolo na seguinte oración: *unha* vs. *catro*:

Os accionistas recibirán unha acción nova gratuíta por cada catro antigas das que certifiquen a súa titularidade. [CORGAetq: CG2000-11-30/8]

O numeral *cardinal* carece formalmente de variación no número; as formas dos *cardinais* que presentan flexión numérica, en realidade, corresponden á acepción de substantivos e adoptan o plural que lles pertence segundo a súa terminación: *ceros*, *uns*,

douses, treses, catros, cincos, etc. Non se deu aínda ningún caso no subcorpus desambiguado pero si se recollen no CORGA:

Xa estaban outra vez cos malditos treses. [CORGA: L. Costas; *Unha estrela no vento*, Xerais, 2000]

Os catros sempre representan a casa, as catro paredes, os catro pés dunha mesa. [CORGA: M. Aneiros; *Da antiga arte de bota-las cartas. A baralla española*, Vía Láctea, 1991]

Polo seu significado, o cardinal *un* refírese só a substantivos singulares mentres que os demais cardinais, posto que indican cantidade superior á unidade, só combinan con substantivos plurais. Esta é a tese que se defende en Álvarez e Xove ou no *Diccionario panhispánico de dudas* (RAE; 2005: 120) cunha pequena diferenza entre eles.

Álvarez e Xove sosteñen que “*os cardinais poden usarse como ordinais e algúns tamén como colectivos, pero son sintacticamente diferentes*”:

Tódolos *un* de maio convocan unha manifestación. Tiven que sentarme na fila *un* [*un* é ordinal: “Tódolos primeiros de maio”, “na primeira fila”]. [Álvarez e Xove, 2002: 520]

A RAE, pola súa banda, aclara o seguinte:

Además de su uso propio para expresar cantidad, los cardinales se emplean a menudo, en el lenguaje corriente, para expresar orden, reemplazando en su función a los ordinales [...] Aunque el uso de los cardinales con valor ordinal es más frecuente cuando se refieren a números altos, también se da en referencia a números bajos; así, se dice tanto *piso siete* como *séptimo piso* o *piso séptimo*. [RAE, 2005: 120]

Álvarez, Regueira e Monteagudo (1986: 269-270) apuntan que non é frecuente que se empreguen os ordinais despois de *décimo* e dan conta de substantivos coma *século, capítulo, páxina*, etc. para os que se emprega indistintamente o ordinal ou o cardinal ata o *décimo*, e a partir de aí o cardinal. Especifican, así mesmo, que “*para designa-lo día do mes úsanse os cardinais (o vinte de xaneiro), agás para o día 1, que pode designarse tamén co ordinal: o primeiro de marzo ~ o un de marzo.*”

Nós cremos que se debe distinguir entre “ser un ordinal” e “expresar valor ordinal”. Creemos que os *cardinais* indican cantidade pero que igualmente poden expresar orde e que esa dualidade debe ser reflectida na etiquetación. Por iso, en XIADA, os *cardinais* poden levar o valor *singular* ou *plural* na súa etiqueta.

Cando indican cantidade, marcámolos como plurais, excepto *un*, *unha*. Así, na oración seguinte, tanto *11* como *catro* etiquétanse Ncdmp (*Numeral, cardinal, determinante, masculino, plural*):

As cifras indicadas no cadro que vai na parte superior dereita desta páxina refírense, como se observa, aos 11 países comprometidos na implantación do euro e deberían ser revisadas coa incorporación dos catro membros restantes. [CORGAetq: *TN1998-05-/18*]

Etiquetamos, non obstante, co número singular cando, por exemplo, se expresan porcentaxes, nas cales o numeral está sempre determinado polo artigo masculino singular; na designación do ano ou do día do mes, co substantivo *aniversario*, *edición*, etc. Nas seguintes oracións a *trinta* e *60* outórgaselles a etiqueta Ncdms (*Numeral, cardinal, determinante, masculino, singular*), etiqueta que tamén reciben, salvo no número que é plural, *7.000* e *90.000* fronte a *oito* e *15* aos que lles corresponde Ncnms (*Numeral, cardinal, non determinante, masculino, singular*):

Esta fin de semana Ortigueira será tomada polos folkis de todo o país e polos incondicionais da festa coa celebración do seu festival por excelencia, que este ano festexa o seu trinta aniversario cun cartel de luxo. [CORGAetq: *ANT2008-07-10/62*]

As vendas de xoguetes descenderon en España durante 1995 uns 7.000 millóns, o que equivale a preto dun oito por cento menos que en anos anteriores, nos que se facturaron 90.000 millóns no sector. [CORGAetq: *CG1996-02-08/105*]

A vicepresidenta adiantara que a intención do Goberno era que o plan se aprobase antes de fin de ano, e onte concretou esa data avanzando a Ban Ki-moon que se fará efectivo na semana en que se conmemorará o 60 aniversario da Declaración dos Dereitos Humanos, que se cumpre o 15 de decembro. [CORGAetq: *GH2008-08-05/58*]

Como acabamos de ver nos exemplos anteriores, a expresión da cantidade represéntase nuns casos mediante letras e noutros con cifras. O mesmo sucede coa expresión da orde, aínda que é moito menos frecuente atopar o *ordinal* escrito en cifra seguido do signo ^o ou ^a. No CORGAetq, versión 1.0, dos 732 numerais ordinais etiquetados, só 2 casos se grafan con cifra. Un deles é o seguinte:

Como se afirmou, Vento mareiro viu a luz en Cuba no ano 1915 (editado o 1º de marzo na Imprenta Artística Concordia y Campanario) e, pouco máis dun lustro despois (o 15 do Samartiño _sic_ do ano 1921), unha segunda edición aparecía en Madrid, dentro da colección "Biblioteca Galicia", incluída na Editorial Galatea e imprentada na casa comercial Juan Pueyo. [CORGAetq: *GH2007-03-17/78*]

No sistema de XIADA, ata o ano 2006, empregabamos a categoría Cifra, sen atributos, para clasificar nela todas as cifras que aparecesen nos textos. Nunha liña semellante creamos ademais a clase Data, tamén sen atributos, para catalogar as distintas aparicións das expresións da data. Este tratamento non é inusual observalo en etiquetarios doutras linguas. Así, para o catalán e castelán, no *tag-set* do corpus do IULA recóllense as categorías Cifra e Data (Morel et al.; 1997: 8; 1998: 9); no do corpus CLIC-TALP para o castelán (Civit et al.; 2001: 22) existe unha clase específica para as Datas e supoñemos, porque non se explicita en ningún momento, que as cifras se integran nos numerais; no do *Corpus Anotado do Português Europeu*³⁸ os algarismos árabes e romanos clasifícanse só como numerais, sen atributos de ningún tipo, etc.

Non obstante, cando comezamos a traballar na etiquetación real de textos, démonos conta da incongruencia que supoñía tratar de distinta maneira os numerais escritos con letras e os grafados con cifras e decidimos eliminar as clases Data e Cifra. Por unha banda, as categorías Cifra e Data non teñen xustificación nas gramáticas; por outro lado, só estabamos postergando a caracterización morfosintáctica deses elementos e, finalmente, toda cifra é susceptible de ser analizada como numeral. Nos seguintes exemplos, fóra a grafía, non hai ningunha diferenza entre o numeral *doce* e a cifra *12*, os dous son numerais e ambos teñen o mesmo referente, polo que a súa etiqueta é idéntica: Ncdmp (*Numeral, cardinal, determinante, masculino, plural*).

Unha muller tivo o DIU "perdido polo corpo" doce anos [CORGAetq: GH2007-03-17/41]

"O que tiña no corpo levaba 12 anos aí metido sen saber que o tiña", explicou Mari Carmen Fernández, que sufría "moitas dores, moitas hemorraxias e infeccións, sen que desen resposta". [CORGAetq: GH2007-03-17/41]

1.3.2.8. Interrogativo-exclamativo (G)

Os atributos que caracterizan esta categoría son:

valor	xénero	número
d (determinante)	m (masculino)	s (singular)
n (non determinante)	f (feminino)	p (plural)
a (determinante/non determinante) ³⁶	a (masculino/feminino)	a (singular/plural)

³⁸ Páxina 3 do manual de anotación, dispoñible en http://www.clul.ul.pt/sectores/linguistica-de-corporus/manual_annotacao.pdf

Exemplos:

Cal *queres?* [Gnas]

Quen *sodes?* [Gnap]

Cantos *quilos precisas?* [Gdmp]

Que *día tan bo!* [Gdms]

Decidimos agrupar nunha única categoría as formas interrogativas e exclamativas *que*, *quen*, *cal(es)*, *canto(s)* e *canta(s)* porque son as mesmas en ambos os dous casos e tanto se empregan para facer unha interrogación parcial directa ou indirecta como se utilizan en exclamacións, con valor principalmente admirativo, “*se ben, dependendo da entoación e do contexto, poden expresar moitos outros matices*” (Álvarez, Regueira e Monteagudo, 1986: 291).

Non incluímos nesta clase os interrogativos *como*, *cando* e *onde* que serán sempre considerados adverbios, tal como recomentan Álvarez, Regueira e Monteagudo (1986: 287).

A normativa sobre a acentuación das formas interrogativas-exclamativas variou no ano 2003, cando xa estaba desenvolvido o etiquetario de XIADA e se aplicaba a fragmentos textuais para probar o preprocesador e o etiquetador. Ata ese ano, prescribíase o acento gráfico en *cal*, *cando*, *canto*, *como*, *onde*, *que* e *quen* cando introducían interrogativas indirectas (RAG/ILG, 1995: 29-30); porén, despois da aparición das novas *Normas* (RAG/ILG, 2003: 31-32), estas formas só se acentúan en casos de anfiboloxía. Este cambio na acentuación dos interrogativos non implicou ningunha modificación no etiquetario, porque non afectou aos atributos e valores caracterizadores, e tampouco alterou o lexicón porque nel xa figuraban tanto as formas con acento como sen el.

Fóra os casos de *cal*, *cales* (variables no número) e de *canto*, *canta*, *cantos* e *cantas* (variables en xénero e número), as demais formas interrogativas-exclamativas caracterízanse por ser invariables. No lexicón a invariabilidade tradúcese en ambigüidade e así *que* e *quen* clasifícanse como Gaaa: *Interrogativo-exclamativo, determinante/non determinante, masculino/feminino, singular/plural*. Na etiquetación

de corpus, sempre que o contexto nolo permite, procuramos desfacer a ambigüidade e asignar o *valor*, *xénero* e *número* que corresponde.

1.3.2.9. Relativo (T)

Os relativos caracterízanse por introduciren unha cláusula subordinada na que cumpren unha función sintáctica, referíndose asemade de maneira anafórica a un elemento ou a un grupo sintáctico anterior no contexto (o antecedente). [Álvarez, Regueira e Monteagudo, 1986: 279]

Consideramos que son relativos as formas *que*, *quen*, *cal*, *cales* e *canto*, *canta*, *cantos*, *cantas*. A estas temos que engadir as formas *cuxo*, *cuxa*, *cuxos* e *cuxas* que introducen as novas *Normas* (RAG/ILG, 2003: 99). Non incluímos, pola contra, na clase Relativo *como*, *cando* e *onde*, os que consideramos adverbios, ben relativos ben interrogativo-exclamativos.

Os atributos que caracterizan esta categoría son:

valor	xénero	número
d (determinante)	m (masculino)	s (singular)
n (non determinante)	f (feminino)	p (plural)
a (determinante/non determinante) ³⁶	a (masculino/feminino)	a (singular/plural)

Exemplos:

Quen chegou tarde foi vostede [Tnas]

A casa na que naceu era moi humilde [Tnfs]

O autor, cuxa obra é ben coñecida, recibiu... [Tdfs]

Polo que respecta ao atributo valor, o relativo *cuxo* é sempre determinante e *que*, *quen*, *cal* son sempre non determinantes. É a forma *canto(s)/canta(s)* a única que pode funcionar como núcleo ou como adxacente:

A educación, pois, debe permitirlles, desde a súa polivalencia, tantas cantas saídas queiran tomar. [CORGA: Ramón Regueira, X. V. Viqueira: *Teoría e praxe*, Novo Século, 1992]

A miña misión é a de marcar goles, cantos máis mellor. [CORGA: GI1996-06/30]

Finalmente, decidimos que as formas *o(s) que*, *a(s) que*, *o(s) cal(es)*, *a(s) cal(es)*, tratadas como formas relativas compostas na *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986: 279-284), no *Esbozo de una nueva gramática de la lengua española*

(RAE, 1973: 218-220) e en Alarcos (1994: 98; 107-109), as iamos analizar en artigo e relativo, respectivamente.

Aínda que como lingüistas nos gustaría identificalas, computacionalmente, a existencia duns casos nos que a unidade *o que* (e variantes) se descompón en artigo e relativo (*O que che dixeron non é certo*) fronte a outras nas que sería unha expresión multipalabra relativa (*Os xogos cos que enreda non son violentos*), que nin sequera é obrigada pois, tal e como sinalan Álvarez, Regueira e Monteagudo (1986: 281) ou Hermida (2004: 95), se o antecedente está expreso non é necesaria a presenza do artigo (*Os xogos con que enreda non son violentos*), faría moi difícil a súa identificación e caracterización computacional. Malia non ser casos semellantes, decidimos caracterizar tamén *o cal* e variantes como dúas unidades: artigo e relativo posto que así nunca errariamos na súa análise.

1.3.2.10. Indefinido (I)

Segundo a *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986: 227), baixo a denominación de Indefinido “*agrúpanse tradicionalmente unha serie de vocábulos caracterizados, en xeral, por daren unha idea de imprecisión, de indeterminación, referida ás nocións de número ou cantidade ou á identidade de algo ou alguén*”.

No noso etiquetario decidimos agrupar cuantificadores e identificadores porque cremos que a distinción é semántica e non funcional. Formalmente, esta é unha clase heteroxénea, pois unhas formas son invariables (*alguén, cada, nada...*), outras varían no número (*tal, bastante*), outras no xénero (*varios, ambos...*) e finalmente, as máis, flexionan xénero e número (*un, algún, ningún, todo, outro...*).

Funcionalmente, tamén se observa moita diversidade entre os membros desta categoría. Hainos que só son nucleares ou non determinantes (*alguén, ninguén, algo, nada, ren, res, un, cadaquén, quenquera*), outros soamente poden funcionar como determinantes (*cada, certo*) e os máis poden desempeñar as dúas funcións (*algún, ningún, todo, outro, moito, pouco, varios, mesmo, propio, tal, tanto, canto, bastante, abondo, demasiado, demais, ambos, entrambos*).

No noso sistema non mantemos a caracterización que Álvarez, Regueira e Monteagudo (1986: 259-261) achegan para *máis* e *menos* por ser contraria, e na nosa opinión non apropiada, á que se proporciona nas *Normas* (RAG/ILG, 1995: 99; 2003: 98) e no *Diccionario da Real Academia Galega*, en diante DRAG, baixo a entrada *máis* (DRAG, 1997: 732). Na *Gramática Galega* inclúen as formas *máis* e *menos* entre as que poden funcionar como nucleares ou como adxacentes e comentan, entre outras particularidades, que estas indican cantidade ou número maior e inferior, e que se empregan nas comparativas de superioridade e inferioridade, respectivamente. No DRAG, esta acepción da *Gramática* categorízase baixo a entrada adverbial e na indefinida recollen unha observación sobre o seu uso: “*Utilízase xeralmente precedido do artigo e seguido da preposición de. Ti fara-lo que fagan os máis. Deixou a máis da comida na mesa*”. Esta descrición concorda coa teoría que nós defendemos e así a aplicamos no sistema de XIADA: *máis* e *menos*, para ser considerados indefinidos, deben levar anteposto o artigo e só poden funcionar como non determinantes. Isto non implica que sempre que vaian precedidos de artigo sexan indefinidos. Nos dous exemplos seguintes constatamos a diferente etiquetación da secuencia *os máis*. No primeiro deles, á forma *máis* correspóndelle a etiqueta Inmp (*Indefinido, non determinante, masculino, plural*) fronte ao segundo que leva a etiqueta Wm (*Adverbio, modificador*):

Os máis dos textos foron escolmados de diferentes publicacións. [CORGAetq: ANT2008-07-10-33]

Os sectores máis castigados son os máis estreitamente ligados ao parón inmobiliario, ao descenso no consumo e ao sector financeiro. [CORGAetq: ANT2008-07-10-72]

Os atributos que caracterizan no etiquetario de XIADA a categoría Indefinido son os seguintes:

valor	xénero	número
d (determinante)	m (masculino)	s (singular)
n (non determinante)	f (feminino)	p (plural)
a (determinante/non determinante)	a (masculino/feminino)	a (singular/plural)
	0 (non aplica)	(non aplica)

Exemplos:

Calquera rapaza [Idfs]

Dáme uns libros calquera [Inmp]

Quero outro calquera [Iams] [Inms]

Non foi ninguén [Inas]

Hai algo aquí que é teu [Inms]

Non me gustou, mais non dixen nada. [In00]

Decidimos empregar por defecto no lexicón o valor ambiguo *masculino/feminino* no xénero, mentres que no número se manteñen os valores individuais *singular* e *plural* na caracterización morfosintáctica dos indefinidos referidos a persoa que sempre son nucleares: *alguén*, *ninguén*, *cadaquén* e *quenquera*; pola contra, nas formas unicamente nucleares que posúen o trazo non humano (*algo*, *nada*, *ren* e *res*) empregamos unha etiquetación dobre: unha, cos valores *masculino singular* (Inms), destinada a caracterizar os casos nos que se explicita esta concordancia e outra, onde non aplica o xénero nin o número (In00), reservada para aqueloutros nos que non se observe ningún tipo de concordancia.

Todos estes elementos indefinidos nucleares marcan, semanticamente, a existencia ou inexistencia de seres e cousas dos que non se coñece o número nin o xénero ou cuxa identidade, cantidade, clase, etc. non interesa precisar. Non obstante, como veremos no capítulo dedicado á etiquetación de corpus, na que o contexto xoga unha baza importante debido a que fai visibles as concordancias entre os distintos elementos, comprobamos que, cando estas formas indefinidas concordan cunha forma verbal, esta está case sempre en singular (exemplos A, B e C fronte a D). Se prestamos atención ás ocorrencias destes indefinidos modificados por un adxectivo variable en xénero, nos referidos a persoa, o xénero aínda que en teoría pode ser masculino ou feminino³⁹, salvo raras excepcións (exemplos de C), aparece en masculino mentres que,

³⁹ Álvarez e Xove (2002: 487), para exemplificar a oposición de xénero nos existenciais de trazo [+hum], proporcionan o seguinte caso: “Para casar, buscaba a *alguén* que fose *rica* e *guapa*”. Nós non atopamos ningún caso de concordancia cun feminino na consulta realizada ao CORGAetq, e nas realizadas contra o CORGA, só detectamos as dúas que incluimos en c) e que nos parecen bastante forzadas.

nos referidos a cousas, só o atopamos en masculino. Os exemplos que amosamos a continuación, extraídos do CORGA, así o demostran:

A) Concordancias coa 3ª persoa singular:

¿Acaso alguén pode discutir que non son de aquí os ourizos máis saborosos? [CORGA: CG2001-10-08/1]

Nada sería peor hoxe que terxiversar. [CORGA: CG1996-02-15/93]

Algo vai mal cando unha sociedade se infantiliza ata estes extremos e chamamos para todo a papa Estado (ou papa Xuíz). [CORGA: GH2007-03-17/86]

B) Concordancias en masculino singular:

Algo parecido ocorreu coa gasolina sen chumbo, que marcou un prezo máximo 87,8 céntimos de euro por litro fronte ó mínimo rexistrado de 73,1 céntimos, o que supón unha diferenza neta sobre o mesmo litro de carburante de ata 14,1 céntimos de euro. [CORGA: CG2002-05-22/10]

Tanto o concello de Pontevedra como a Asociación pola Defensa da Ría non outorgan credibilidade a estas informacións e a oficina de prensa de Celulosas de Pontevedra limitouse a dicir que "non hai nada novo", pero sen confirmar nin desmentir a noticia. [CORGA: ANT2002-10-10/25]

A igualdade vende e é tan sinxela de conseguir como quedou demostrado, pero de verdade alguén serio e cabal pode pensar que unha lei de número, de repartición de escanos pola metade _na marxe doutras disposicións menos evidentes_, lles pode traer xustiza ós méritos da muller. [CORGA: GH2007-03-17/81]

Pola outra, si no sexo, eisí, en absoluto, ren é malo e ren é anómalo, tódalas experiencias, tódalas aventuras e tódalas imaxinacións, sin límite algún, deben estar permitidas. [CORGA: D. García Sabell, *As facianas do erotismo cotemporáneo*, Galaxia, 1976]

C) Concordancias en feminino singular:

Pois mira, tiñan razón, eu fun a túa querida, e ben querida e, que saiban todo, que nunca ninguén foi mellor querida por alguén. [CORGA: Vicente Araguas, *A canción do verán*, Ir Indo, 2001]

Daquela, nunca é límite (como proporía o individualismo) nin obxecto para ser utilizado, senón alguén coa que medra a persoa ó establecer unha relación dialóxica. [CORGA: X. M. Domínguez Prieto, *Alén da libido. do sexo á sexuidade*, Espiral Maior, 1998]

D) Concordancias coa 3ª persoa plural:

Ninguén sospeitabamos que as medidas a tomar foran xa as tan mortuorias dos nichos de Bonaval. [CORGA: CG1999-05-21/7]

Entrevistado: A Fraga , por ventura reconvertido á democracia, hai que lembrarlle un pasado ó que ninguén podemos renunciar. [CORGA: X. A. Liñares Giraut, *Conversas con Avelino Pousa Antelo. Memorias dun Galego Inconformista*, Edicións do Castro, 1991]

1.3.2.11. Pronome (R)

Como xa sinalamos con anterioridade, en XIADA a clase Pronome está constituída só polo tipo tradicionalmente denominado *persoal* cuxo paradigma “*posúe distintas formas segundo a función que poidan desempeñar dentro do enunciado. Agora ben, tódolos pronomes persoais se caracterizan pola súa función nuclear dentro da frase nominal e por non necesitaren determinantes*” (Álvarez, Regueira e Monteagudo, 1986: 157-158). Esta categoría, ademais, caracterízase por ser a única na que se conservan restos da antiga declinación latina, pois só nela se mantén a oposición de caso ao ofrecer formas diferenciadas para o *nominativo*, *acusativo*, *dativo* e *preposicional*. Así, por exemplo, o paradigma do pronome para a segunda persoa singular presenta *ti* para o caso nominativo; *te* para o acusativo; *che* para o dativo e, finalmente, *ti* e *contigo*, no preposicional.

En galego moderno caeu en desuso a forma tónica neutra da terceira persoa, *elo*, empregada no galego antigo. Hoxe en día, tal como se afirma en Álvarez, Regueira e Monteagudo (1986: 161) ou, máis recentemente, en Fernández Salgado (2004: 461) baixo a entrada *elo*, unicamente aparece formando parte dalgunhas expresións fixas como *estar a elo* ('estar realizando algo con afán'), sacado que nunca coa función de suxeito. Álvarez e Xove (2002: 536) corroboran a existencia na lingua antiga do pronome *elo*, mais manifestan que existiron formas diferenciadas para o nominativo (*elo*) e o complemento (*nelo*, *delo*). No que si están todos de acordo é en que os seus usos primitivos foron asumidos polo pronome masculino *el* e polos demostrativos neutros *isto*, *iso* e *aquilo*.

A realidade lingüística actual obríganos a matizar a anterior afirmación. Nós non negamos a desaparición da antiga forma *elo* no galego actual, (é máis, como galegofalantes nativos nunca a empregamos), pero nos textos actuais atopámonos con frecuencia cunha unidade *elo* cuxa aparición se debe, case sempre, á plena vixencia do neutro castelán *ello*, que provoca que, por interferencia, se empregue tamén en galego tanto no caso preposicional como no nominativo. A normativa actual recomenda limitar

o seu uso ás expresións galegas xenuínas e non empregalo na función de suxeito porque “a posibilidade de funcionar como suxeito probablemente non a desenvolveu nunca” (RAG/ILG, 2003: 90).

No etiquetario de XIADA reflíctese o uso que, por calco do castelán, se fai desta forma na lingua escrita actual e por iso a mantemos no lexicón no paradigma pronominal coa etiqueta Rtr3ns (*Pronome, tónico, nominativo/preposicional, terceira, neutro, singular*). Os seguintes exemplos constitúen unha proba da súa utilización:

O ordenador, por poñer un exemplo, non é (aínda) universal, malia que pouco lle pode faltar para elo, e o seu uso resulta imposible en circunstancias onde o medio escrito posúe unha maior disponibilidad, dada a súa independencia de fontes enerxéticas. [CORGA: P1997-04/4]

E elo sen ter que resolver previamente, na onda de ascenso na que se sitúa desde 1993, dous dos seus déficits: [CORGA: TN1997-11/13]

Cando cremos que o saco estaba limpo, procedemos a elo apertando contra o costado da chalana. [CORGA: X. Rodríguez Vergara; *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

Os atributos cos que caracterizamos a categoría Pronome en XIADA son os seguintes:

tipo	caso	Persoa	xénero	número
a (átono)	a (acusativo)	1 (primeira)	m (masculino)	s (singular)
t (tónico)	d (dativo)	2 (segunda)	f (feminino)	p (plural)
	n (nominativo)	3 (terceira)	n (neutro)	a (singular/plural)
	p (preposicional)		a (masculino/feminino)	
	f (formante léxico)			
	r (nominativo/preposicional) ⁴⁰			
	o (outros casos)			

⁴⁰ Este valor só se emprega para caracterizar algúns pronomes no lexicón (*ti*, por exemplo). Na etiquetación de corpus, a ambigüidade expresada polo valor “r” desfáise en dous valores distintos, un para *nominativo* e outro para *preposicional*; o contexto indicaranos en cada caso cal corresponde.

Exemplos:*se* [Rao3aa]*Faille caso a túa nai!* [Rad3fs]*Vés comigo?* [Rtp1as]*Se puidese, comíate a bicos.* [Raa2as]*Din que llo impide a súa conciencia.* [Rad3as lle] [Raa3ms o]

Como se pode deducir da táboa e dos exemplos anteriores, si respectamos a caracterización das formas pronominais segundo a súa tonicidade e así recoñecemos dúas series, unha tónica e outra átona; non obstante, non seguimos as recomendacións de EAGLES respecto a distinguir no *tipo* do Pronome as formas dos pronomes reflexivos e recíprocos nin empregamos o atributo *cortesía*. Sobre este último véxase o apartado 1.3.2.2., centrado na categoría Verbo, onde se explica por que non diferenciamos entre formas verbais cortesés e non cortesés e cuxa argumentación é totalmente apropiada para a clase Pronome.

Polo que se refire a discriminar no tipo entre pronomes *persoais*, *reflexivos* e *recíprocos*, cremos que, por unha banda, sería incoherente posto que o termo *persoal* integra xa os outros dous subtipos. É dicir, cando pensamos no paradigma pronominal, a primeira distinción que se nos ocorre é a de separalos en tónicos e átonos e a continuación clasificalos segundo o caso ou función sintáctica que poidan desempeñar, e non en formas reflexivas e non reflexivas.

Na proposta de EAGLES, por outro lado, a reflexividade aplícase só as formas átonas e non se ten en conta para as tónicas. Amais, en galego a única persoa que posúe formas diferenciadas para expresar a reflexividade é a terceira: *si* e *consigo* na serie tónica e *se* na átona. Nós entendemos que cos atributos *tipo*, *caso* e *persoa* podemos caracterizar morfolóxica e sintacticamente os distintos pronomes sen necesidade de incorporar a diferenciación entre formas *reflexivas* e *non reflexivas* pois ademais:

1) as formas átonas dos reflexivos (*me*, *te*, *se*, *nos*, *vos*, *se*) coinciden formalmente coas dos recíprocos (*me*, *te*, *se*, *nos*, *vos*, *se*) e só tendo en conta a forma verbal á que acompañan, sendo aínda así frecuentes os enunciados equívocos, podemos decantarnos por un ou outro valor;

2) morfosintacticamente, as formas tónicas reflexivas *si* e *consigo* caracterízanse sempre no caso co valor *preposicional*; pola contra, as formas átonas poden remitirse ao caso *acusativo* (pois case sempre desempeñan a función de C.D.), *dativo* (inusualmente poden funcionar como C.I.) ou incluso non desempeñar función sintáctica ningunha (verbos tipo *arrepentirse*):

O pronome reflexivo ten a función de complemento directo e non son posibles en galego construcións reflexivas nas que a acción se exerza unicamente sobre unha parte do suxeito: *lavei as mans, queimeei un dedo* e non **laveime as mans, queimeime un dedo*, nas que o pronome reflexivo sería complemento indirecto. (...) Unicamente pode darse un dativo reflexivo cando equivale a *para si* ou en construcións de reciprocidade: *Resérvome o dereito de interpelalo* (ou, preferiblemente, *reservo para min...*); *Déronse moitos golpes*; son frecuentes os casos con verbos de lingua: *Pregúntome a miúdo que será dela; Dixenme que non podía ser e nada máis; Dixéronse cousas tremendas un ó outro*". [Álvarez, Regueira e Monteagudo, 1985: 179]

Hai unha serie de verbos, chamados pronominais, que no estado actual da lingua só se poden contruír coa forma reflexiva. Neles a idea dun obxecto que recibe a acción verbal está moi debilitada e en ningún caso admiten un complemento directo distinto do propio suxeito, polo que están moi próximos aos verbos intransitivos, mesmo se podería dicir que o pronome actúa como morfema intransitivizador [Álvarez, Regueira e Monteagudo, 1985: 181]

Así pois, en XIADA as formas *me, te, nos* e *vos*, reflexivas ou non, son etiquetadas no lexicón cos valores casuais *acusativo, dativo* e *formante léxico*. Este último valor habilítamolo para clasificar o pronome átono reflexivo co que se constrúen verbos como *arrepentirse, atreverse, inmutarse*, etc., o cal non desempeña función sintáctica ningunha na oración ou, como postula Cidrás (1991: 113-114), non é índice verbal de ningunha función sintáctica.

Será nunha fase posterior, en concreto na de desambiguación, onde se lle asignará a cada unha destas formas (*me, te, nos* e *vos*) a etiqueta que corresponda, como nos seguintes exemplos:

1) *Formante léxico*:

Refirome ó de face-la declaración da renda por teléfono. [CORGAetq: CG1996-02-11/5]

Werner non é só un firme partidario do 'euro' senón tamén da súa introducción en xaneiro de 1999, porque "o día no que tivese que ser aprazada a Unión Europea Monetaria (UEM),

temería outra terrible presión á alza para o marco alemán, e entón poderíamos esquecernos totalmente de todo pacto pro emprego. [CORGAetq: CG1996-02-29/102]

"Os luns, mércores e venres penso que debo irme; [CORGAetq: CG2002-05-10/6]

2) *Acusativo*:

Ás veces albisco o remate deste túnel fraguista no que o país está sumido e síntome aliviado coma un cadelo ao que acaban de quitarlle un carracho. [CORGAetq: ANT2005-03-03/6]

"Os veciños tratáronnos de marabilla", destacan os afectados. [CORGAetq: GH2007-03-17/6]

Cando buscas o que te une e non o que te separa, chegas a acordos positivos. [CORGAetq: ANT2005-03-03/19]

3) *Dativo*:

Aínda hoxe, cando recordo o "bilingüismo harmónico" prodúceme un alérxico arrepío. [CORGAetq: GH2007-03-17/82]

Pregúntome que pensa a xente cando le informaci3ns como as que aquí reflectimos todos os días. [CORGAetq: GH2007-03-17/17]

Mayka Brañas: _"Eu 3s dous meses de comezar na serie leveime un desgusto tremendo 3 saber que estaba embarazada de tres meses. [CORGA: GH2005-02-17/60]

Non incluímos na descri33n anterior o pronome átono reflexivo de terceira, *se*, porque este, á parte dos valores *acusativo*, *dativo* e *formante léxico* que pode expresar calquera das denominadas formas reflexivas átonas, presenta tamén, como valores tradicionais característicos, a marca de impersonalidade e de pasiva reflexa, ou na terminoloxía empregada por Frías e López (2001: 127-142), *impersoal intransitiva* e *impersoal transitiva*.

Sen pretender meternos en demasiadas fonduras nin en revisar as distintas teorías que existen sobre os distintos valores de *se*⁴¹, recolleamos, cos seguintes exemplos extraídos do CORGA, unha mostra de cada un dos valores xeralmente recoñecidos para esta forma:

Barbeouse, lavouse, vestiuse, e logo dun cacho volveu coas súas estrañezas: [CORGA: X. López López; *O caderno*, 2001] [**reflexivo acusativo**]

⁴¹ Neste sentido consideramos fundamental para o galego o estudo de Cidrás Escáneo (1991: 101-125) "As construcións pronominais en galego. Proposta de clasificación e caracterización".

"China resérvase o dereito de toma-las medidas que considere oportunas se a OTAN, liderada por Estados Unidos, non cumpre o esixido polo noso país", engadiu. [CORGA: CG1999-05-14/17] [**reflexivo dativo**]

Sempre cando xurdían problemas, e xurdiron moitos estando as cousas como estaban, se protexeran un ó outro. [CORGA: U. Heinze; *Remuíños en coiro*, 1984] [**recíproco acusativo**]

Os novos amigos despedíronse dentro do bus e, antes de separarse, el deulle un beixo en cada meixela mentres lle apertaba a man como se lle estivese pedindo socorro. Nos días seguintes, escribíronse un par de cartas e concertaron unha cita. [CORGA: X. Fuentes Castro; *O pescador de cana*, 2006] [**recíproco dativo**]

Nela, o edil pedía para o concello o regalo dos reis para que lle trouxesen os cartos necesarios para embelecer todo o posible o seu municipio e estamos seguros de que, máis tarde ou máis cedo, conseguirá todo o que se propón posto que, como el mesmo di na súa misiva, foi moi bo e traballador e o concello merecese un regaloño. [CORGA: CG1995-02-08/143] [**dativo ético ou valor expresivo**]

Nada máis dicilo arrepeníuse, pero xa era tarde. [CORGA: M. Lourenzo González; *Arqueofaxia*, 1995] [**formante léxico**]

El levántase cedo todos os domingos e vai tocar as campás. [CORGA: M. Rivas; *¿Que me queres, amor?*, 1996] [**voz media ou intransitivizador**]

Trabállase en varias posibles prospectivas. [CORGA: CG1996-02-08/7] [**impersoal, con verbo intransitivo**]

E ése máis rico ensoñando neles que ténndoos gardados no peto. [CORGA: R. Vidal Bolaño; *Touporroutou da Lúa e do Sol (Farsada choqueira para actores e bonecos, ou viceversa)*, 1996] [**impersoal, con verbo copulativo**]

Por conseguinte, a identidade galega como realidade obxectiva será unha construción social e persoal, sempre dinámica e estruturada consonte coa seguinte dialéctica: a) de socialización nunha tradición e nunha lingua: internalización desa cultura. b) de proxección e creación persoal a partir dos alicerces nos que se foi inculturado e, finalmente, c) de obxectivación do creado persoal e comunitariamente. [CORGA: CG2000-11-26/59] [**impersoal, con verbo en pasiva**]

De feito, fanse spots publicitarios que non cumpren a legalidade. [CORGA: ANT2008-06-26/11] [**pasiva reflexa, con verbo en plural e suxeito gramatical posposto**]

O feito era que, como dicía a mai do Tarrío, "por moito que o demo cave, todo se sabe". [CORGA: X. Alcalá; *Latitude austral*, 1991] [**pasiva reflexa, con verbo en singular e suxeito gramatical anteposto**]

Coa anterior exemplificación pretendemos facer reparar nas dificultades que un lingüista atoparía para clasificar o pronome *se* co obxectivo de xustificar a etiqueta única que lle outorgamos no sistema de XIADA: Rao3aa (*Pronome, átono, outros casos, terceira, masculino/feminino, singular/plural*).

No deseño inicial do etiquetario, como nos sucedeu con outras categorías, fomos demasiado ambiciosos e pretendemos caracterizar os distintos valores da forma *se*. Así, no atributo *caso*, ademais dos que figuran na táboa xeral, incluíamos o valor *impersoal* (i) e o xenérico ambiguo *acusativo/dativo/impersoal* (b)⁴². Cando comezamos a revisar e corrixir a etiquetación automática realizada sobre un subcorpus de textos xornalísticos do ámbito da economía⁴³, démonos conta da complexidade do problema pois, ás dificultades que tiñamos para identificar os diferentes valores de *se*, debíamos sumar a tarefa de facer posible que estes fosen recoñecidos por un etiquetador automático. Nese momento cuestionámonos o etiquetario e decidimos modificalo creando un valor novo, *outros casos*, que caracterizase sempre, única e exclusivamente, a forma *se*.

A solución pola que optamos non é innovadora; os etiquetarios do castelán e do catalán do corpus do IULA consideran no atributo correspondente ao tipo de pronome o valor “6” para o que especifican “*passiu pronominal-impersonal / personal, feble, cas pendent, tercera persona, nombre pendent, gènere pendent*”. E na explicación do “pronom R6” aclaran:

Amb el codi R6 es pretén donar compte del funcionament de les formes catalanes *es, s', 's, se* i la castellana *se* en tant que passives pronominals-impersonals i pronominals personals febles de tercera persona amb cas (acusatiu o datiu), gènere i nombre pendents d'especificar. La primera possibilitat (passives-impersonals) descriu exemples com els següents: *se lavan coches, s'esdevingué tot de sobte, se subastan cuadros*. La segona (pronominals personals febles de tercera persona) en recull els següents: *La Maria es pentina, se bebió la leche, se situen en dues direccions*. [Morel et al., 1998: 17-18]

⁴² Véxase Domínguez (2008).

⁴³ O corpus de adestramento empregado en XIADA é posible consultalo [<http://corpus.cirp.es/corgaetq>] ou ben descargalo [<http://corpus.cirp.es/xiada/descargas.html>]. A versión descargable á que nos referimos é a 2.3 (con 304.118 elementos analizados) e consta de varios arquivos en formato texto que permiten ver a etiquetación realizada sobre un conxunto de noticias xornalísticas completas do ámbito temático da economía, extraídas aleatoriamente do CORGA (corpus_economia_prensa); todas as noticias das distintas seccións do xornal *Galicia Hoxe* do 17 de marzo do 2007 e do número correspondente ao 3 de marzo do 2005 do semanario *A Nosa Terra*, extraídas tamén do CORGA (corpus_xeral_prensa) e finalmente, oracións de diferentes xéneros e procedencia cuxo obxectivo é incluír todas as etiquetas posibles do etiquetario de Xiada polo menos unha vez (nucleo).

O etiquetario do Seminario de Lingüística Informática (SLI) da Universidade de Vigo (Aguirre Moreno et al., 2002b: 221) non proporciona tampouco máis ca unha caracterización xenérica para *se*, á que lle corresponden as etiquetas e definicións seguintes:

PPS3AR: pronome persoal átono singular terceira reflexivo

PPS2CAR: pronome persoal átono singular segunda de cortesía reflexivo

PPP3AR: pronome persoal átono plural terceira reflexivo

PPP2CAR: pronome persoal átono plural segunda de cortesía reflexivo

No *Corpus Anotado do Portugués Europeu*⁴⁴, por poñer un exemplo máis de indefinición na caracterización morfosintáctica de *se*, todos os pronomes persoais átonos son remitidos á categoría Clítico (CL), para a cal non consideraron pertinente establecer ningún atributo; *se*, polo tanto, igual ca todas as demais formas átonas, unicamente recibe a etiqueta de clítico.

No etiquetario de XIADA, as formas pronominais tónicas son clasificadas con respecto ao atributo *caso* en nominativas e preposicionais. *Min, comigo, contigo, si, consigo, connosco* e *convosco* remítense sempre ao caso preposicional. *Ti, vostede(s), el(es), ela(s), nós, vós, nosoutros(as)* e *vosoutros(as)* son ambiguas, pois tanto poden aparecer en funcións correspondentes ao caso *nominativo* como ao *preposicional*. Pola contra, *eu* é o único pronome cuxo valor casual é sempre o *nominativo*, aínda que na lingua actual atopámolo tamén nalgunha ocorrencia, lingüisticamente incorrecta, onde debería aparecer a forma *min*, correspondente ao caso *preposicional*:

Cando o vin entrar non souben que facer, porque levaba case tres anos sen ver a máis ninguén ca eu e esquencérame do que se facía cando un batía con outra persoa. [CORGA: Miguel-Anxo Murado; *Memoria de derribos*, Galaxia, 1993.

Entrevistado: _Son un indisciplinado, non proba os papeis que fago, non ensaio e o elenco artístico que me acompaña é tan independente e bohemio coma eu mesmo. [CORGA: CG1996-02-10/179]

No paradigma pronominal tónico, o xénero está explícito só na terceira persoa do singular *el(es), ela(s), elo* e nos plurais inclusivos de primeira e segunda plural, respectivamente, *nosoutros(as)* e *vosoutros(as)*. Nos demais elementos, caracterizados

⁴⁴ Páxina 3 do manual de anotación, dispoñible en http://www.clul.ul.pt/sectores/linguistica-de-corpus/manual_annotacao.pdf

no lexicón co valor *a* (*masculino/feminino*), tal como veremos no capítulo dedicado á desambiguación, só poderemos desfacer a ambigüidade naquelas ocasións en que o xénero do pronome se explicita mediante concordancias cos seus modificadores, como sucede nos casos seguintes:

Nó-las dúas marchamos dirixiuse Sheila á porta de saída levando da man á compañeira. [CORGA: M. Suárez Abel; *Selva negra*, Xerais, 1994]

"De pequena gustábame reciclar trapos para facer os traxes e vestidos que eu mesma inventaba" [CORGA: *DLV2007-05-11/2*]

Polo que respecta ao número, parécenos interesante salientar a oposición que se dá en galego “*entre as formas nós / vós, inclusivas porque fan referencia a un plural que inclúe os interlocutores e mais un grupo que non ten ningunha limitación; e nosoutros(as) / vosoutros(as), exclusivas porque se refíren a un grupo pechado e ben delimitado*” (Hermida, 2004, 74). Ex.:

Nosoutras, pensadoras de finais do século XX, sabemos (...) que o pensamento non é a reprodución da realidade, senón o mostrarse da realidade a través dun sistema historicamente determinado de mediacións, in primis a mediación lingüística. [CORGA: C. Blanco; *O contradiscurso das mulleres. (Historia do proceso feminista*, Nigra, 1995)

Na serie átona, normalmente distínguese entre formas de acusativo (*me, te, o/lo/no, nos, vos, os/los/nos*) e formas de dativo (*me, che, lle, nos, vos, lles*). A oposición de xénero e número só se manifesta mediante morfema gramatical na terceira persoa, sendo esta persoa ademais, xunto coa segunda singular, a única que posúe formas diferenciadas para o caso *acusativo* e o *dativo*:

		Singular	Plural
Acusativo	Masculino	o, lo, no	os, los, nos
	Feminino	a, la, na	as, las, nas
Dativo		lle	lles

O pronome acusativo de terceira persoa, tanto singular como plural, presenta distintos alomorfos en distribución complementaria:

- *no, na, nos* e *nas* empréganse cando a forma verbal remata en ditongo. Exs.:

O seu director xeral, José Luis Losada, atribuíuno ós "esforzos de xestión". [CORGAetq: *CG1996-02-15/34*]

Problemas, hainos. [CORGAetq: *TN1997-04/7*]

- *lo, la, los e las* úsanse despois de forma verbal acabada en *-r* ou *-s*, tras os dativos *lles e nos* e despois do pronome interrogativo *u*:

Aumenta-los anos de cotización e atrasa-la xubilación, dúas propostas para palialo [CORGAetq: *CG1996-02-16/105*]

¿Víchela en algures? [CORGA: A. Risco; *Tánata*, Xerais, 1998]

¿Ulas as delicias do matrimonio que tanto lle pregoaran? [CORGA: M. X. Queizán; *A semellanza*, Sotelo Blanco, 1992]

Os restantes detalles deixámoslles ós propios moradores, así téñense en que ocupar durante anos interminables. [CORGA: U. Heinze; *Remuíños en coiro*, Xerais, 1984]

Tampouco alas, porque nolas cortaron... [CORGA: X. Rábade Paredes; *Como levar un morto*; Sotelo Branco, 2001]

- *o, a, os e as* utilízanse nos demais casos:

Segundo Solbes, esta marxe "nin a temos agora, nin a teremos no futuro". [CORGAetq: *CG1996-02-14/109*]

Aplique o lector oído polas rúas e prazas deste país e comprobarao. [CORGAetq: *GH2007-03-17/82*]

A énclise do pronome átono de primeira plural, *nos*, independentemente do seu valor casuístico (*acusativo, dativo* ou *formante léxico*), provoca nas formas verbais de primeira persoa plural a perda do *-s* final do morfema de número e persoa:

"Cada vez que chove botámonos a tremer, e Urbaser levou todo antes de que chegasen as cámaras". [CORGAetq: *GH2007-03-17/6*]

Contámonolo todo. [CORGA: X. A. Neira Cruz; *As cousas claras*, Xerais, 2001]

A historia demóstranos que detrás das intervencións humanitarias sempre hai intereses económicos privados que terminan provocando unha crise maior en beneficio propio, lembrémonos de Iraq. [CORGAetq: *GH2008-08-05/97*]

Na determinación dun etiquetario, as peculiaridades que acabamos de ver para o pronome de terceira *acusativo* ou o de primeira plural non revisten maior importancia, posto que non afectan á etiqueta que debe clasificalos. É na elaboración da gramática formal destinada a facilitar a incorporación das unidades léxicas ao lexicón, no recoñecemento morfosintáctico das distintas unidades e na reconstrución e xeración de

formas onde deberemos habilitar os mecanismos ou regras necesarios para lidar coas numerosas complicacións que provocan nun sistema automático estes trazos.

No galego estándar debe observarse a distinción no dativo de terceira persoa entre unha forma singular *lle* e outra plural *lles*: Rad3as (*Pronome, átono, dativo, masculino/feminino, singular*) fronte a Rad3ap (*Pronome, átono, dativo, masculino/feminino, plural*). Con todo, nos textos actuais aínda en ocasións se reflicte o descoñecemento da forma plural *lles* (Álvarez, Regueira e Monteagudo, 1986: 164) ao constatar a utilización de *lle* para singular e para plural; en consecuencia, a forma *lle*, maioritariamente singular, é etiquetada en casos coma os seguintes co valor plural⁴⁵:

Habitualmente están dispostos a participar de xeito desinteresado en proxectos que lle gusten e sexan serios. [CORGAetq: ANT2005-03-03/72]

Este centro, ademais de achegarlle formación ás persoas con discapacidade, terá tamén unha área de formación para profesionais e os medios necesarios para a utilización das novas tecnoloxías. [CORGAetq: GH2007-03-17/32]

Na serie átona aséntase aínda outra distinción. Trátase da diferenciación de funcións que se establece para a segunda persoa singular, entre unha forma de caso *dativo, che, e* outra para o valor casual *acusativo e formante léxico, te*. As seguintes oracións exemplifican, respectivamente os tres *casos*:

Agora ninguén se fai rico cun disco, pero un traballo discográfico sérveche de escusa para dar concertos e que te coñezan". [CORGAetq: ANT2005-03-03/69]

Por exemplo, se ti vas a un hotel facilmente te atoparás cun cadro moderno, non figurativo... é algo normal. [CORGAetq: ANT2005-03-03/49]

Finalmente, para rematar coa descrición das formas do paradigma pronominal galego, só nos queda dar conta da súa representación na escrita e das contraccións que con elas se producen.

Álvarez, Regueira e Monteagudo sintetizan claramente o primeiro destes aspectos:

Os pronomes átonos enclíticos ortografíaanse unidos directamente ó verbo: *escachároncheme a xerra, dixéronllelo todo*. O conxunto formado polo verbo e o(s) pronome(s) forma un todo a

⁴⁵ Na versión 1.0 do CORGAetq, consultado o 27/04/2010, que consta de 304.118 elementos gramaticais analizados, das 437 ocorrencias do pronome átono de terceira dativo, singular ou plural, en 19 ocasións dáse a utilización de *lle* por *lles*, o que supón unha frecuencia do 4,67%. Describiremos con máis detemento estes usos no capítulo dedicado á desambiguación.

efectos de acentuación: *deron mais déroncheme ben ganas de rir* [Álvarez, Regueira e Monteagudo, 1986: 165]

Se deixamos a un lado as contraccións, podemos afirmar que a ortografía dos pronomes átonos non repercute na súa etiquetación. Provoca iso si, como veremos máis adiante, múltiples problemas á hora de introducir os verbos no lexicón, de recompoñer a forma verbal cando esta, por levar algún pronome enclítico, sofre modificacións ben acentuais ben de desaparición de caracteres, etc.

Polo que respecta ás contraccións nas que están implicados algúns pronomes, debemos distinguir entre as que se producen cos da serie tónica e as que teñen lugar cos da serie átona, pois as primeiras afectan a elementos de dúas categorías gramaticais distintas mentres que nas segundas só concorre unha.

Os pronomes tónicos de terceira persoa, *el, ela, elo, eles* e *elas* contraen coas preposicións *en* e *de* dando lugar ás seguintes formas: *nel, nela, nello, neles, nelas; del, dela, delo, deles, delas*. Exemplos:

O 75 por cento deles remataron en sentenza condenatoria por parte dos tribunais. [CORGAetq: CG1996-02-09/95]

Nelas traballan unhas 300.000 persoas. [CORGAetq: GH2007-03-17/39]

No corpus de adestramento do etiquetador automático non figura ningunha contracción co neutro *elo*, pero estas si aparecen na versión 1.4 do CORGA, malia que cunha frecuencia moito menor ca calquera das outras contraccións de pronomes tónicos. Así, fronte ás 3481 ocorrencias de *deles* ou as 429 de *nelas, delo* concorre en 89 ocasións e *nello* faino en 26, coa particularidade de que 47 casos de *delo* e 15 de *nello* rexístranse nunha única obra⁴⁶.

Os pronomes átonos de dativo *me, che, lle, nos, vos* e *lles* cando concorren co pronome da terceira persoa acusativo dan lugar ás seguintes contraccións:

	me	che	lle	nos	vos	lles
o/lo	mo	cho	llo	nolo	volos	llelos
a/la	ma	cha	lla	nola	vola	llela
os/los	mos	chos	llos	nolos	volos	llelos
as/las	mas	chas	llas	nolas	volas	llelas

⁴⁶ Conde, Alfredo (1995): *Sempre me matan*, Vigo, Ir Indo.

O tratamento que reciben todas as contraccións é idéntico: a unidade gráfica é segmentada en tantos constituíntes como elementos posúe a contracción, e, a continuación, cada un deles recibe a etiqueta que lle corresponde. No caso dos pronomes tónicos, polo simple feito de estar ante unha contracción, xa sabemos de antemán que non vai haber ambigüidade en ningún atributo; o xénero e o número están explicitados nas distintas formas e o caso, á forza, é preposicional. Por exemplo, se tomamos a contracción *del*, esta constitúe unha unidade que posúe dous constituíntes, o primeiro deles a *preposición de* e o segundo o *pronome tónico, preposicional, terceira, masculino, singular*:

Análise unidade ▾
Unidade
del
Constituín ▾
Forma de Etiqueta P Lema de
Constituín ▾
Forma el Etiqueta Rtp3ms Lema el

Nas contraccións dos átonos, a ambigüidade inicial no atributo *caso* que se dá nas formas *me*, *nos* e *vos* queda tamén resolta, posto que para que poida ter lugar a contracción deben funcionar como dativas. Agora ben, formalmente establécese unha diferenza nestas contraccións entre o alomorfo do pronome de acusativo que contrae coas formas de dativo singulares (*me*, *che*, *lle*) e entre a variante que contrae coas dativas plurais: *o* vs. *lo*.

Se nos fixamos nas dúas análises seguintes, observamos como as unidades léxicas *llo* e *nolo* se desagregan en dous constituíntes; o primeiro deles recupera en cada unidade a *forma* completa que lle é propia: *lle* (*Pronome, átono, dativo, terceira, masculino/feminino, singular*) e *nos* (*Pronome, átono, dativo, primeira, masculino/feminino, plural*), respectivamente. A parte común ás dúas unidades é a etiqueta e o lema do segundo constituín, *o* (*Pronome, átono, acusativo, terceira, masculino, singular*); mentres que o diferencial asoma na *forma*: *o*, variante empregada con forma verbal que non remate en ditongo nin en *-r* ou *-s* e tras os pronomes *me*, *che* e *lle* fronte a *lo*, variante utilizada con forma verbal rematada en *-r* ou *-s* e tras os pronomes *nos*, *vos* e *lles*:

Análise unidade ▾ Unidade Ilo
Constituínte ▾ Forma lle Etiqueta Rad3as Lema lle
Constituínte ▾ Forma o Etiqueta Raa3ms Lema o

Análise unidade ▾ Unidade nolo
Constituínte ▾ Forma nos Etiqueta Rad1ap Lema nos
Constituínte ▾ Forma lo Etiqueta Raa3ms Lema o

1.3.2.12. Adverbio (W)

O único atributo que caracteriza esta categoría en XIADA é o tipo:

tipo
n (nuclear)
m (modificador)
a (nuclear/modificador) ⁴⁷
r (relativo)
g (interrogativo-exclamativo)

Exemplos:

Terriblemente [Wa]

A noticia afectoume terriblemente [Wn]

Estaba terriblemente incómoda [Wm]

Pregúntalle cando chega [Wg]

A aldea onde me criei chámase Larín [Wr]

O adverbio caracterízase por ser un tipo de palabra invariable en xénero e número. Tradicionalmente, segundo a natureza da circunstancia expresada, os adverbios clasifícanse en adverbios de lugar, tempo, modo, cantidade e precisión (Álvarez,

⁴⁷ O valor *nuclear/modificador* só caracteriza o adverbio no lexicón; na etiquetación de corpus o valor “a” desagrégase en dúas etiquetas, unha co valor *nuclear* e outra co valor *modificador*; o contexto indicará cal corresponde.

Regueira e Monteagudo, 1986: 421-422). A estes, as *Normas* (RAG/ILG, 1995: 173-178; 2003: 173-179) suman os de afirmación, negación e dúbida e Costa Casas et al. (1988: 229-253) engaden os inclusivos e os especificadores de orde, realce e límite.

No noso etiquetario, aínda que nun principio pensamos en realizar unha delimitación semántica do adverbio, decidímonos finalmente por unha clasificación morfosintáctica porque cremos que será de máis utilidade. Unha das características do adverbio é que pode modificar un verbo, un adxectivo, outro adverbio ou mesmo unha oración enteira pero el non pode ser modificado máis que por outro adverbio ou unha frase preposicional. Así pois, consideramos pertinente aplicar nesta categoría o atributo *valor* tendo en conta que con el distinguimos entre *nuclear* e *modificador*. Nas seguintes oracións, os adverbios *onte* e *hoxe* etiquétanse como Wn (*Adverbio, nuclear*) mentres que *ben* é etiquetado como Wm (*Adverbio, modificador*):

Os datos foron facilitados onte polos responsables da 35 edición da Feira Internacional do Xoguete, Hobby e Carnaval (FEJU), que abre hoxe as súas portas ata o próximo día 12. [CORGAetq: CG1996-02-08/105]

Galicia tamén sae ben parada polas estatísticas. [CORGAetq: CG1999-05-06/3]

Tamén pensamos, e nisto seguimos a Álvarez, Regueira e Monteagudo (1986: 285-286), que *como*, *cando* e *onde* son sempre adverbios, ben relativos ben interrogativos-exclamativos.

Así nas oracións seguintes *cando* e *onde* reciben a etiqueta Wr (*Adverbio, relativo*):

Xa non nos vale a escusa de hai dez ou quince anos, cando en Santiago sempre chovía. [CORGAetq: ANT2005-03-03/25]

En toda empresa hoxendía, pequena, mediana ou grande, hai un lugar, próximo á papeleira, onde se depositan os currículos que van chegando. [CORGAetq: ANT2005-03-03/4]

Ás formas *como*, *cando* e *onde* interrogativas-exclamativas, debemos incorporar tamén as variantes acentuadas *cómo*, *cándo* e *ónde*. Tal como se viu xa na categoría do Interrogativo-exclamativo, dende a reforma ortográfica do 2003 as formas acentuadas deben empregarse unicamente en casos de anfiboloxía, malia que nos textos non sempre se respectan as normas establecidas e podemos atopalas sen que exista ambigüidade.

Nas seguintes oracións recóllese algún exemplo de adverbios interrogativos-exclamativos, anterior e posterior a 2003:

O chanceler alemán, Helmut Kohl, declarábase onte, sábado, canso de discusións sobre cómo, cándo e con quen comezará a unión económica e monetaria europea (UEM), e recomendoulles a tódolos países da UE que fagan os seus deberes. [CORGAetq: CG1996-02-25/81]

O asesoramento necesítase para saber cando vender e acadar o mellor prezo. [CORGAetq: ANT2002-10-17/29]

Costa-Gavras, coa súa película Le Couperet, acaba de resaltar de forma dramática cómo humilla, aniquila, deprime e até pode volver a un tolo a perda do traballo. [CORGAetq: ANT2005-03-03/53]

A mesma fonte tamén aclarou que aínda é pronto para saber onde situarán uns e onde se colocarán outros. [CORGAetq: GH2007-03-17/73]

Na proposta de EAGLES recoméndase caracterizar o adverbio co atributo *grao* e distinguir deste xeito entre formas adverbiais positivas, comparativas e superlativas. Nós non incluimos no noso etiquetario este atributo porque só uns poucos adverbios admiten gradación e, se o introducísemos, por coherencia tamén deberíamos caracterizar dalgún xeito aqueloutros que admiten o sufixo alterativo *-iño* ou as diferenzas formais que se establecen nos adverbios rematados en *-mente* cando estes se coordinan.

Na *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986) resúmese o problema da clasificación do adverbio cunha reflexión moi esclarecedora na que entran as peculiaridades anteriores:

Como se ve, esta é unha clasificación puramente semántica, que pervive aínda nas gramáticas por mor da dificultade de encadrar con criterios morfolóxicos e sintácticos esta categoría heteroxénea de palabras. Proba desta heteroxeneidade é que algúns adverbios admiten gradación (*máis preto, mellor, cedísimo*) pero outros non (**máis onte, *semprísimo, *axiñísima...*); algúns intensifican o seu significado usando o sufixo *-iño* (*a modiño, cerquiña*), mentres que outros non poden: **nunquiña, *acolaíña, etc.*; e finalmente, algúns deles (os de lugar, os de tempo e algún outro) poden ir precedidos de preposición (*por aquí, para cerca, de lonxe, ata despois, desde agora, etc.*), en cambio os máis non o admiten: **de paseniño, *para así, *por abondo, etc.* [Álvarez, Regueira e Monteagudo, 1986: 422]

No etiquetario de XIADA non se teñen en conta estas particularidades e, en consecuencia, excepto os relativos e os interrogativos-exclamativos, que levan unha etiqueta específica (Wr, Wg), e mais as formas apocopadas *tan* e *moi*, que sempre son

modificadoras (Wm), todos os demais adverbios recibirán por defecto no lexicón a etiqueta Wa (*Adverbio, nuclear/modificador*), quedando o *valor* sen determinar ata a fase seguinte, a de etiquetación e desambiguación. Non obstante, isto non implica que no sistema de XIADA non se recoñezan os adverbios cuxas formas presenten un sufixo de grao, un alterativo, etc. É máis, xa tivemos que enfrontarnos con estas cuestións á hora de construír o corpus de adestramento, posto que aparecen nos textos seleccionados para este fin, porén o seu tratamento verémolo no capítulo seguinte. Unha mostra constitúena os exemplos que ofrecemos a continuación nos que as formas *moitísimo*, *pretiño* e *porcentual* reciben a etiqueta Wn (*Adverbio, nuclear*):

España está case na dianteira, pero tamén en Italia ou en Alemania hai xente nova cun alto nivel; en Los Ángeles hai un grupo de profesores excelente; en China un nivel técnico alucinante porque traballan moitísimo... [CORGAetq: ANT2005-03-03/49]

Vivimos na rúa Predicadores (como xa creo ter dito neste Diario de auga) pretiño dos restos das murallas romanas, do río Ebro pola altura da Ponte de Santiago (a ponte máis ancha de España, cando a ergueron, en 1968), e a dous inmobles de distancia da tapería El Papeo, onde eu, a diario, papeo, claro, un petisco, contra as 14.30 horas. [CORGAetq: GH2008-08-05/54]

Os estranxeiros con permiso de residencia en España non están distribuídos porcentual nin territorialmente. [CORGAetq: ANT2008-07-10/7]

1.3.2.13. Conxunción (C)

Na actualidade só clasificamos as conxuncións en coordinantes ou subordinantes: *c* (coordinante), *s* (subordinante). Porén, inicialmente fomos moi ambiciosos co tratamento das conxuncións. A simple subdivisión en coordinantes e subordinantes parecíanos escasa e decidimos completala incluíndo o subtipo de cada unha delas. Así, ata finais de 1999, os atributos considerados foron:

Conxunción	
tipo	subtipo
c (coordinante)	c (copulativa)
s (subordinante)	d (disxuntiva)
	a (adversativa)
	cc (concesiva)
	cd (condicional)
	ca (causal)
	cs (consecutiva)
	f (final)
	l (locativa)
	t (temporal)
	m (modal)
	cm (comparativa)
	cp (completiva)

Cando comprobamos que era inviable aplicar nun proceso automático tal grao de descrición, posto que os resultados da etiquetación automática eran desastrosos, decidimos quedarnos só co atributo tipo e eliminar o subtipo que clasificaba as distintas conxuncións.

Non tivemos en conta os atributos subtipo coordinante e subtipo subordinante de EAGLES por considerar que a información que proporcionaban e os criterios que os delimitaban pertencían ao nivel sintáctico.

Se lembramos a súa proposta, comprobamos que non se presentaba unha clasificación semántico-sintáctica do subtipo, senón que os criterios de delimitación dos subtipos de conxuncións coordinantes non tiñan nada que ver cos valores propostos para os subtipos subordinantes:

(ii)	Coord-Type:	1. Simple	2. Correlative	3. Initial	4. Non-initial
(iii)	Subord.-type:	1. With-finite	2. With-infin.	3. Comparative	

Así, mentres o criterio clasificador nas conxuncións coordinantes é de tipo posicional:

This attribute subclassifies coordinating conjunctions. It is easier to assign one tag to one orthographic word and it is therefore suggested that the four values are assigned as follows:

Simple applies to the regular type of coordinator occurring between conjuncts: German *und*, for example. When the same word is also placed before the first conjunct, as in French *ou...ou...*, the former occurrence is given the **Correlative** value and the latter the **Simple** value. When two distinct words occur, as in German *weder...noch...*, then the first is given the **Initial** value and the second the **Non-initial** value. [Leech e Wilson, 1996]

nas subordinantes, tal como se deduce dos valores establecidos, está directamente relacionado coa forma verbal que segue á conxunción ou co tipo de subordinada que introduce:

Further information on the Type of Subordination, i.e. if the conjunction requires a finite verb *+fin*, a non-finite verb *+infv* or introduces a comparison *compar* is proposed on Level 2b, since such phenomena are shared by three different languages. [Monachini e Calzolari, 1996]

Neste momento no lexicón de XIADA figuran as seguintes conxuncións:

Coordinantes: *ben, e, mais, nen, nin, ora, ou e quer.*

Subordinantes: *unque, apenas, así, asemade, ca, cal, cando, coma, como, conforme, conque, consonte, daquela, emporeso, emporiso, entanto, entón, entonces, logo, mais, mentras, mentres, namentres, pero, pois, porén, porque, que, se, segundo, senón e xa.*

Como se pode observar, os nexos adversativos recóllense nos subordinantes debido á súa proximidade cos concesivos, figurando só nos coordinantes os copulativos e mais os disxuntivos ou distributivos.

Todas estas conxuncións son formas normativas excepto *nen, entanto, mentras*, e *unque*; as tres primeiras nunca o foron, pero a última sufriu un cambio de estatus nas últimas *Normas* (RAG/ILG, 2003), pasando de normativa a non normativa. O motivo último da súa presenza no lexicón de XIADA, como o de todos os elementos nel incluídos, é facilitar o recoñecemento morfolóxico dos documentos do CORGA, no que os textos incorporados reflicten a non completa consecución da normativización do galego na actualidade.

1.3.2.14. Locución (L)

Unha locución é unha combinación fixa de palabras que funciona como unha unidade de distinto tipo (conxuntiva, adverbial, preposicional, substantiva, adxectiva ou verbal) cuxo significado non é a suma dos elementos que a compoñen. Locución non é

obviamente unha nova categoría gramatical, senón un xeito de indicar que se corresponde cunha unidade multipalabra.

No etiquetario de XIADA só previmos a existencia das adverbiais, prepositivas e conxuntivas. Non consideramos as substantivas, adxectivas e verbais por unha razón moi sinxela. Nas primeiras non hai flexión de ningún tipo, son invariables fronte ao substantivo, adxectivo e verbo que son categorías nas que a variación é enorme (xénero, número, tempo, modo, etc.). Incluir as locucións substantivas, adxectivas e verbais implicaría integrar baixo unha mesma clase (locución) categorías que responden a valores moi diversos e que non son pertinentes para definir todos os elementos.

O motivo principal que nos levou á consideración da clase Locución radicou en que, coa súa existencia –fronte ao adverbio, preposición e conxunción–, facilitamos a extracción dos elementos así catalogados para desenvolver no futuro un necesario estudo deste conxunto de unidades multipalabra.

A categoría Locución agrupa, polo tanto, as conxuncións, preposicións e adverbios que están constituídos por unidades multipalabra. Os atributos tipo e subtipo permiten a súa identificación e caracterización:

Tipo	Subtipo
a (adverbial)	c (coordinante)
p (preposicional)	s (subordinante)
c (conxuntiva)	0 (non aplica)

Exemplos:

a continuación, de balde, ... [La0]

a través de, con respecto a, ... [Lp0]

e mais, ou ben [Lcc]

aínda que, de non ser que, ... [Lcs]

Como xa vimos, na proposta de EAGLES recoméndase integrar as unidades multipalabra nas correspondentes categorías gramaticais simples, aínda que se permite tamén tratalas dun xeito diferente (Monachini e Calzolari, 1996). O fundamental é sempre que a información poida ser extraída e homoxeneizada coa doutros etiquetarios

ou aplicacións e nós, neste aspecto, non teriamos ningún problema en recategorizar os elementos clasificados na clase Locución nas correspondentes simples, posto que a información está contida no etiquetario, só que a un nivel distinto.

1.3.2.15. Preposición (P)

O grupo EAGLES denomina esta clase *Adposición* e nela considera pertinentes os seguintes atributos:

L0	Adposición
L1	Tipo: preposición, posposición, circumposición
	Formación: simple, contracción
	Xénero: masculino, feminino, neutro
	Número: singular, plural

Os subtipos *posposición* e *circumposición* son específicos para linguas como o inglés (os usos do denominado *xenitivo saxón*) ou o alemán:

German *entlang* is a **Postposition**, and arguably, the 's which forms the genitive in English is no longer a case marking, but an enclitic postposition, as in *the Secretary of State's decision, in a month or so's time*. German (*auf...*) *hin* is an example which can be analysed as a **Circumposition**. [Leech e Wilson, 1996]

Obviamente, nin no castelán, portugués, catalán ou galego tería razón de ser esta distinción, polo que non entendemos que no etiquetario do Corpus CLiC-TALP se manteña o subtipo *preposición* na categoría Adposición/Preposición⁴⁸.

Á parte do *tipo*, a determinación da Preposición mediante os atributos *formación*, *xénero* e *número* semella cando menos contraditoria.

Por unha banda, se en algo están de acordo as gramáticas é en que as preposicións son invariables en canto á forma e, como tales elementos invariables, non son pertinentes os atributos xénero ou número. Estes poden empregarse para caracterizar o artigo co que contrae a preposición, en linguas como o castelán ou o

⁴⁸ En Civit et al. (2001: 21-28) fálase da categoría Preposición, pero no exemplo de corpus etiquetado e desambiguado, *ante* ou *de*, por exemplo, aparecen coa etiqueta SPS00, que, como explica a propia Civit (2003: 120), correspóndese coa categoría *Adposición*, tipo *preposición*, formación *simple* onde non aplica nin o xénero nin o número por estar estes dous atributos reservados para as contraccións.

portugués, se se decide non considerar as contraccións como categorías especiais⁴⁹ ou se non se establecen etiquetas separadas para os constituíntes das contraccións tal como se recomenda encarecidamente dende EAGLES.

Por outra banda, ao ver o atributo *formación*, debemos recoñecer que entre os seus valores esperabamos, ao lado de *simple*, o de *locución* e non o de *contracción*. Sería lóxico que co termo *formación* se acollese a diferenza entre a unidade simple e a multipalabra pois “*nuns casos trátase dunha soa palabra; noutros, de grupos de palabras (normalmente sustantivos ou adverbios seguidos dunha preposición simple), coñecidos como ‘locucións prepositivas’, fixadas xa pola lingua e en nada distintas en canto á función con respecto ás ‘preposicións simples’*” (Álvarez, Regueira e Monteagudo, 1986: 477).

Diciamos sería lóxico porque ambos os valores caracterizarían a mesma categoría gramatical: a de preposición. Non obstante, atopámonos con que co atributo *formación* e o valor “*fused*” se pretende darlles cabida ás contraccións de preposición e artigo:

The additional value **Fused prep-art** is for the benefit of those who do not find it practical to split fused words such as French *au* (= *à* + *le*) into two textwords. This very common phenomenon of a fused preposition + article in West European languages should preferably, however, be handled by assigning two tags to the same orthographic word (one for the preposition and one for the article). [Leech e Wilson, 1996]

Na nosa opinión, a solución proposta non parece demasiado acertada. É incongruente, primeiro, por admitir no interior dunha categoría gramatical (Preposición) un subtipo, como mínimo, bicategorial (Preposición e Artigo). Logo, deixando á parte a cuestión da amálgama de clíticos, por moito que en castelán, catalán ou italiano as contraccións máis frecuentes ou únicas sexan as de preposición e artigo, en portugués, sen ir máis lonxe, as preposicións *de* ou *em* contraen co artigo definido, co indefinido, co pronome persoal tónico de terceira ou cos demostrativos. Pois ben, dende EAGLES non se prevén as contraccións de preposición con outra clase de palabras máis ca co artigo. Así, en Leech e Wilson (1996) as contraccións de preposición con outro elemento que non sexa artigo non se teñen en conta, e en Monachini e Calzolari (1996),

⁴⁹ No nivel morfolóxico, o modelo de EUROTRA para o castelán asigna *al* e *del* a unha categoría combinada. Así mesmo, o modelo multilingüe de GENELEX para o portugués trata as contraccións de preposición e determinante como unha unidade morfolóxica especial.

malia que se propugna o desmembramento de calquera tipo de contracción tanto para as categorías como para as etiquetas implicadas, encontrámonos co subtipo *contracción* só nas clases Preposición e Residual. Para esta última, ademais, non se aclara absolutamente nada nos comentarios xerais (só se recolle o valor na táboa categorial) e tampouco se menciona nin aparece na descrición das aplicacións ás distintas linguas de traballo da proposta (portugués, castelán, italiano, catalán, alemán, francés, grego, irlandés ou holandés).

Aínda que máis adiante, nun apartado propio, trataremos con detalle o tratamento que reciben as numerosas contraccións do galego, podemos adiantar aquí que van ser desagregadas en tantos elementos como constituíntes posúan, caracterizando individualmente cada un deles coa etiqueta e categoría que lles corresponda.

Volvendo ao tema obxecto desta epígrafe, a clase Preposición en XIADA non presenta atributos e os elementos nela clasificados son os seguintes:

a	cas	entre	sacando
aga	con	excepto	salvante
agás	conforme	fôra	salvo
ante	consoante	malia	según
antre	consonte	mediante	segundo
após	contra	menos	sen
até	de	onda	senón
ata	deica	para	sin
baixo	dende	perante	so
bardante	des	por	sobor
cabo	desde	pra	sobre
canda	durante	quitando	tras
cara	en	sacado	xunta

Desta relación dedúcense tres feitos:

1) Non figura ningunha preposición multipalabra. Estas, como vimos no apartado anterior, trátanse no noso sistema na clase das locucións.

2) Recóllense elementos desbotados ou non recomendados pola normativa vixente: *antre*, *cara*, *pra*, *según*, *sin* e *sobor*. As preposicións *pra* e *sobor* aparecen

recollidas no *Vocabulario Ortográfico da Lingua Galega* (VOLGa) (González e Santamarina, 2004) como formas toleradas mais non aconsellables, recomendando no seu lugar empregar *para* e *sobre*; *antre*, *cara*, *según* e *sin* utilízanse no CORGA, motivo polo que se inclúen na listaxe. Dada a complexidade da cuestión e as súas repercusións, non imos abordar neste momento a necesidade de dar cabida a unidades léxicas non normativas; baste saber que estas serán caracterizadas e analizadas tanto no lexicón como no corpus.

3) A procedencia dos elementos en xeral, como veremos no capítulo centrado no deseño e construción do lexicón, e das preposicións en particular é diversa. Nalgunha ocasión, para as clases pechadas, a fonte son as *Normas* (RAG/ILG, 2003)⁵⁰ (*após*); na maioría dos casos, o VOLGa (todas as preposicións excepto *após*, *antre*, *cara*, *según* e *sin* teñen como fonte este vocabulario) e os distintos dicionarios do galego actual documentan a orixe do léxico de XIADA, pero o común denominador de todos eles, polo menos nas categorías cun número de constituíntes limitado, é a súa presenza no CORGA. Esa é a razón de non incluír entre as preposicións o elemento *catar*, recollido no VOLGa e nos dicionarios de Ir Indo (Ledo Cabido, 2004) ou Xerais (Carballeira Anllo, 2000), por exemplo, pero non documentado no CORGA. Nunha consulta realizada na versión 1.4 do CORGA⁵¹ que consta duns 25 millóns de palabras, das 77 ocorrencias da forma *catar* ningunha era preposición. Por curiosidade, realizamos a consulta tamén no *Corpus Técnico Anotado do Galego* (CTAG)⁵², no *Corpus Técnico do Galego* (CTG)⁵³, no *Corpus Lingüístico da Universidade de Vigo* (CLUVI)⁵⁴ e no *Tesouro Informatizado da Lingua Galega* (TILG)⁵⁵ con resultados dispares: os dous primeiros non amosaron ningún resultado; o corpus do CLUVI recolle un caso correspondente ao verbo e no TILG, das 217 ocorrencias da palabra *catar*, só unha está etiquetada como preposición⁵⁶, a pesar de que se nos guiamos polas *Normas* (RAG/ILG,

⁵⁰ *Após* lístase na relación de preposicións das *Normas* (RAG/ILG, 2003: 180), pero non no VOLGa.

⁵¹ Dispoñible en <http://corpus.cirp.es/corga> [consultado o 31/03/2010]

⁵² Dispoñible en <http://sli.uvigo.es/CTAG/> [consultado o 05/04/2010]

⁵³ Dispoñible en <http://sli.uvigo.es/CTG/> [consultado o 05/04/2010]

⁵⁴ Dispoñible en <http://sli.uvigo.es/CLUVI/> [consultado o 05/04/2010]

⁵⁵ <http://www.ti.usc.es/TILG/> [consultado o 31/03/2010]

⁵⁶ Documentada na tese de doutoramento *Textos dialectais do norte da Terra Chá* de Xosé Luís Regueira Fernández (1989): "*¡Oh! Chegaban lonxe. E eso que daquela non había coches nin... Catar se había bicicletas.*"

1995: 184; 2003: 185) non poderíamos delimitala nas preposicións, senón que habería que considerala locución conxuntiva condicional.

1.3.2.16. Interxección (Y)

Na proposta de EAGLES non se prevé ningún trazo para caracterizar a categoría Interxección e tampouco se empregan outros atributos nas aplicacións descritas polo grupo para as linguas italiana, alemá, inglesa, francesa, grega, sueca, catalá ou irlandesa. O danés e o portugués non tratan esta clase e o castelán non aparece recollido nas distintas aplicacións. En consecuencia, tampouco no etiquetario de XIADA temos en conta ningún atributo.

A listaxe de interxeccións que están integradas no lexicón de XIADA é a seguinte:

ó	cataplún	hurra	pataplaf
abur	chino	isca	pataplún
adeus	churra	madía	plof
ah	clac	madeus	puf
ai	clic	malia	pum
ala	coriño	malpocado	pumba
ale	crac	mícaro	quino
aleluia	dodemo	múcaro	recoiro
alerta	ea	miau	salve
alto	eh	mis	socorro
amén	ei	miso	tintín
arre	eiramá	oh	to
avante	epa	oi	toma
ave	fo	olé	uf
bah	fou	ola	upa
benia	gache	opa	ux
boh	glogló	ora	vaia
bueno	ha	ou	vaites
ca	hala	ouh	veña
canté	ho	oxalá	viva
carafío	hola	paf	xo
caramba	hum	pam	zas

1.3.2.17. Sinais de puntuación (Q)

En Monachini e Calzolari non se inclúen na táboa das clases de palabras nin se describen posteriormente os sinais de puntuación, só se menciona de pasada a súa utilidade en relación coa diferenciación entre descrición léxica e etiquetario de corpus:

Sometimes tag classes are in reality different from lexical descriptions. For example, classes for punctuation are needed and certain types of semantic or pragmatic or lexical information can be present in the tags (e.g. the days of the week). [Monachini e calzolari, 1996]

É en Leech e Wilson (1996) onde se propoñen dúas maneiras de categorizalos:

- Na primeira das opcións cada signo de puntuación implica un atributo *valor* único que o caracteriza: punto, coma, etc.
- Na segunda das opcións, agrúpanse en subclases dependendo da súa posición na oración. Así hai signos de puntuación de final de oración (. ? !), intraoracionais (, ; : -) e sinais de puntuación de unidades parentéticas iniciais (([) ou finais ()]).

En XIADA decidímonos pola primeira opción e os sinais de puntuación que recoñecemos son os seguintes:

<i>Sinal de puntuación</i>	<i>Etiqueta</i>
(guión baixo)	Q
- (guión curto)	Q-
, (coma)	Q,
; (punto e coma)	Q;
: (dous puntos)	Q:
! (pecha admiración)	Q!
? (pecha interrogación)	Q?
/ (barra diagonal)	Q/
. (punto)	Q.
... (puntos suspensivos)	Q...
' (comiña simple)	Q'
" (comiñas dobres)	Q"
((abre paréntese)	Q(
) (pecha paréntese)	Q)
[(abre corchete)	Q[
] (pecha corchete)	Q]
{ (abre chave)	Q{
} (pecha chave)	Q}
¡ (abre admiración)	Q¡
¿ (pecha admiración)	Q¿

1.3.2.18. Categoría periférica (Z)

Coa implantación desta categoría pretendemos darlles cabida aos moitos elementos residuais que aparecen nos distintos textos e que non podemos remitir a ningunha das grandes clases de palabras. Normalmente os gramáticos considéranos fóra do seu ámbito de estudo e en consecuencia non os describen e case ningún os recolle sequera. Así, as abreviacións, fórmulas ou símbolos non se tratan en Álvarez, Regueira e Monteagudo (1986), nin en Costa Casas et al. (1988), nin en Álvarez e Xove (2002). Pola contra, Cunha e Cintra (1984: 116-117) recoñecen a existencia das abreviacións e achegan unha lista das siglas máis coñecidas, mentres que o *Esbozo de una nueva gramática de la lengua española* (RAE, 1973: 154-159), pola súa banda, ofrece unha listaxe das abreviaturas máis comúns. Ambas as obras, non obstante, limítanse a constatar a existencia de abreviacións e a facilitar pequenas listas; ningunha delas delimita categorialmente nin profunda na súa análise.

O grupo EAGLES, consciente da necesidade de incluír unha clase de palabras á que remitir os elementos xeralmente non tratados nas gramáticas pero frecuentes nos textos, recomenda a creación da categoría Residual:

The residual value (R) is assigned to classes of textword which lie outside the traditionally accepted range of grammatical classes, although they occur quite commonly in many texts and very commonly in some. For example: foreign words, or mathematical formulae. It can be argued that these are on the fringes of the grammar or lexicon of the language in which the text is written. Nevertheless, they need to be tagged. [Leech e Wilson, 1996]

Os atributos que consideramos pertinentes no noso etiquetario para caracterizar os elementos incluídos nesta clase son os seguintes:

tipo	xénero	número
a (abreviatura)	m (masculino)	s (singular)
f (fórmula)	f (feminino)	p (plural)
g (sigla)	a (masculino/feminino)	a (singular/plural)
s (símbolo)	0 (non aplica)	0 (non aplica)
o (outros tipos)		

No bienio 2005-2006, cando afrontamos a reorganización do lexicón, decidimos tamén eliminar da Categoría periférica o subtipo *estranxeirismo* e volver caracterizar os

elementos así clasificados recategorizándoos; é dicir, as palabras pertencentes a linguas distintas do galego que, ben dun xeito consciente ben dunha maneira inconsciente, se empreguen nos textos a procesar etiquetaranse como calquera outra palabra do galego e remitiranse á categoría gramatical que lles corresponda. Impórtanos máis a súa caracterización morfosintáctica que o feito de que sexa unha forma allea á nosa lingua. Por exemplo, nas secuencias seguintes as unidades *prime time*, *sex simbol* ou *estrés*, anglicismos en diferentes estados de adaptación ao galego, compórtanse como calquera outro substantivo no referido a determinantes, concordancias, etc. O mesmo sucede coa preposición latina anglicizada *versus*⁵⁷ cuxo comportamento é idéntico ao da súa homóloga galega *contra*.

Destaca o prime time, onde se emiten dúas series de ficción, Como no cine e O Clon. [CORGAetq: GH2007-03-17/88]

De feito, a sex simbol que marcou os anos 60 e 70, alcumada O Corpo, non era senón Raquelita Tejada, filla dun inmigrante boliviano e emparentada coa que foi presidenta de Bolivia Lidia Gueiler Tejada. [CORGAetq: ANT2008-07-10/54]

Nun mundo que está dominado polo estrés as benzodiazepinas axudanlle a xente a relaxarse e tamén lle permiten durmir con moita mellor calidade. [CORGAetq: ANT2008-07-10/11]

O G-5 versus o G-8 [CORGAetq: ANT2008-07-10/79]

Decidimos discriminar os distintos tipos de abreviacións porque a súa formación, flexión de xénero e número, representación na escrita e lectura difiren dun a outro. Malia que os atributos xa estaban establecidos, González Rei (2004) foinos de grande axuda para asentarmos os criterios de clasificación e obter listaxes amplas das abreviaturas, siglas e símbolos.

As abreviaturas fórmanse, mediante apócope ou síncope, eliminando parte das letras ou sílabas dunha palabra ou dun sintagma; os atributos de *xénero* e *número* son pertinentes para a caracterización morfolóxica daquelas abreviaturas en cuxas correspondentes formas plenas tamén son necesarios; rematan sempre cun punto ou cunha barra diagonal e lense como a palabra plena que abrevian. Exemplos:

Termino Sres. Accionistas. [CORGAetq: CG1996-02-07/65]

Sres. Zamp [Categoría periférica, abreviatura, masculino, plural, (lema) Sr.]

⁵⁷ Así cualificada no *Diccionario panhispánico de dudas* da RAE, [entrada] *versus*, p. 666.

De seguida o director de administración, Sr. Couñago, da lectura ó Balance e a Conta de Perdas e Ganancias ó 31/12/96. [CORGAetq: F1997-08/2]

Sr. Zams [Categoría periférica, abreviatura, masculino, singular, (lema) Sr.]

Por exemplo, en Galicia supón ao redor de 75.000 millóns de ptas.: [CORGAetq: G11996-05/55]

ptas. Zafp [Categoría periférica, abreviatura, feminino, plural, (lema) pta.]

En todo caso, alguén debería aclararlle á cidadanía que menores impostos significan menores e/ou piores servicios públicos: seguridade, educación, saúde, etc. [CORGAetq: ANT2002-11-07/29]

etc. Za00 [Categoría periférica, abreviatura, xénero non aplica, número non aplica, (lema) etc.]

As siglas, en xeral, parafraseando a González Rei (2004: 274-279), fórmanse coas iniciais dos substantivos e adxectivos que conforman o sintagma ou ben coas iniciais de todas as palabras do sintagma; formalmente son invariables en xénero e número aínda que estes en numerosas ocasións maniféstanse a través das concordancias; escríbense sen puntos e en maiúsculas, e a súa lectura depende da propia sigla: se se trata de siglas silábicas lense coma calquera outra palabra; se son consonánticas deletréanse, e noutras ocasións combínanse ambos os sistemas mesturando a silabación co deletereo. Vexamos algún exemplo de etiquetación de siglas:

Tamén hai recursos da UE para o traslado de empresas por razóns ecolóxicas. [CORGAetq: ANT2005-03-03/19]

UE Zgfs [Categoría periférica, sigla, feminino, singular, (lema) UE]

En calquera caso, os efectos da supresión do IAE son certos e inmediatos e supoñen un beneficio electoral para o Partido Popular. [CORGAetq: ANT2002-10-17/30]

IAE Zgms [Categoría periférica, sigla, masculino, singular, (lema) IAE]

O Plan xeral de control tributario para 1999 profundará nas actuacións para detecta-las tramas de fraude organizada, segundo publicaba onte o Boletín Oficial do Estado (BOE). [CORGAetq: CG1999-05-26/47]

BOE Zg00 [Categoría periférica, sigla, xénero non aplica, número non aplica, (lema) BOE]

Non diferenciamos entre siglas e acrónimos, pois estes, a pesar de diferir na formación (poden tomar varias letras dunha, de varias ou de todas as palabras que o conforman) e na ortografía (só con inicial maiúscula, en xeral), compórtanse morfoloxicamente coma as siglas. As siglas ou os acrónimos que xa se consideran

palabras comúns, pois escríbense con minúscula e ata teñen entrada propia nos dicionarios, considerámoslos substantivos. Vexamos algún exemplo de cada:

En total o presuposto do Sergas para este exercicio acada os 2.205,7 millóns de euros (367.005 millóns de pesetas), o que supón un 6,10% de incremento respecto ó 2001. [CORGAetq: CG2002-05-12/2]

Sergas Zgms [*Categoría periférica, sigla, masculino, singular, (lema) Sergas*]

IBM lanzou hai dous anos outro proxecto similar para axudar científicos do Instituto de Investigación Scripps, en La Jolla (California), a analizar compostos químicos para atopar unha cura contra a sida. [CORGAetq: GH2007-03-17/60]

sida Scfs [*Substantivo, común, feminino, singular, (lema) sida*]

Rodeado de bits, vive a contragolpe de quen se sabe aldeán de orixe e galáctico de destino. [CORGAetq: GI1996-06/30]

bits Scmp [*Substantivo, común, masculino, plural, (lema) bit*]

Os símbolos, segundo González Rei (2004: 281) son “abreviacións establecidas por convención utilizadas fundamentalmente no mundo da ciencia e da técnica e instituídas por organismos internacionais competentes no campo en que se empregan, co fin de facilitar o intercambio de información”.

A diferenza das abreviaturas, os símbolos son formas fixas internacionais; aínda que o seu desenvolvemento ou lectura se realice de xeito distinto en cada lingua, o símbolo é coñecido e interpretado internacionalmente do mesmo xeito. Por exemplo, os símbolos que representan os elementos químicos, os elementos matemáticos, as unidades do Sistema Internacional de Unidades (SI), os puntos cardinais, etc.

Caracterízanse tamén, fronte ás abreviaturas, por ser invariables posto que non admiten morfemas de xénero ou número e ademais non levan punto nin acento. En XIADA son sempre caracterizados só co subtipo *símbolo*; é dicir, neles non aplican os atributos de *xénero* e *número*. No CORGAetq o máis empregado, con diferenza, é o símbolo da porcentaxe seguido do símbolo “menos”. Vexamos algúns exemplos reais de uso:

A finais do verán a realidade impoñía que o crecemento deste ano non superará o 2% aínda que o goberno central insiste en que poderá estar polo 2,2%. [CORGAetq: ANT2002-10-03/32]

Ademais, non variaron os prezos dos produtos de medicamentos e comunicacións, mentres que descenderon os de lecer e cultura (-1,3 por cento) e os de hoteis, cafés e restaurantes (-0,7 por cento). [CORGAetq: CG2001-10-12/1]

Os populares cren que deberían figurar entre os proxectos prioritarios da UE o desdoblamento do gasoduto troncal e a súa conexión coa Meseta; o novo eixe de conexión de Alta Tensión en 400 KV entre Galicia e Asturias; a instalación do segundo circuito de interconexión con Portugal; e un novo eixe de 400 KV que transforme o actual de 220 KV en Mesón-Belesar-Montearenas. [CORGAetq: GH2007-03-30/14]

O subtipo *fórmula* está previsto para caracterizar, sen que aplique nel ningún atributo, as distintas fórmulas (matemáticas, químicas, etc.) que poidan aparecer nos documentos. No corpus de adestramento do etiquetador de XIADA non apareceu ningunha fórmula, malia que estas si están presentes en textos que polo momento non foron etiquetados⁵⁸. Por exemplo, en obras correspondentes tematicamente ao bloque das Ciencias e a Tecnoloxía, atopamos entre outros casos os seguintes:

neste proceso, $Q=m.Lv$, e teremos así o noso valor para LLv . [CORGA: R. Vilalta et al., *Física e química*, A Nosa Terra, 1997]

Supoñamos na superficie da Terra unha presión $P = 101\ 300$ Pa. [CORGA: R. Vilalta et al., *Física e química*, A Nosa Terra, 1997]

A fórmula empírica de moitos glúcidos _que contén o número total de átomos_ é $C_x(H_2O)_y$, e por iso antigamente aos glúcidos chamóuselhes hidratos de carbono ou carbohidratos, parecía que o carbono estaba enlazado a unha molécula de auga; [CORGA: R. Varela Díaz, *Biología. Segundo Curso de Bacharelato*, Bahía, 1999]

Logo, para $G = 1999,18$ teremos, aplicando a fórmula proposta, que $M = 1419,89$; [CORGA: R. Vilalta López, *Astronomía práctica e outras cousas*, Toxosoutos, 1999]

Deriva de aplicar a seguinte fórmula: $IMC (Kg/m^2) = peso (Kg) / [talle (m) \times talle (m)]$ [CORGA: C2007-05/25]

O subtipo *outros*, no que non aplican tampouco os atributos *xénero* e *número*, integra todos aqueles elementos que non podemos remitir a ningunha outra clase de palabra e os cales non sabemos como clasificar. En xeral, a ese *caixón* van os enderezos

⁵⁸ Polo momento, na versión de consulta web do CORGA non están representados os caracteres que aparecen como subíndices ou superíndices, aínda que si están marcados nos documentos procesados. Así, por exemplo, para o primeiro caso mostrado, a oración codificada no CORGA contén a marcaxe do subíndice: "neste proceso, $Q=m.L<subíndice>v</subíndice>$, e teremos así o noso valor para $LL<subíndice>v</subíndice>$."

web, as matrículas de vehículos, os números de teléfono, os identificadores como a), A) ou 1._, ou os prefixos e sufixos. Exemplos:

Podes consultar o cartel na web www.klubbers.com ou no teléfono [902499994](tel:902499994). [CORGAetq: GH2007-03-30/113]

Respecto do accidente rexistrado na Pobra, J.G.B.A., de 47 anos e veciño de Santander, faleceu ao chocar o turismo que conducía, un BMW matrícula [7147BRB](#), contra un camión Volvo _matrícula [4484CPT](#)_. [CORGAetq: GH2008-08-05/24]

b) Conseguir o necesario control sobre o noso sistema económico, tanto nos aspectos produtivos como nos financeiros. [CORGAetq: MC1980-12/1]

Outra forma de incluír nunha palabra o significado de pé é utilizar no composto o prefixo ou sufixo grego podo (podómetro, podólogo, antípoda, monipodio, artrópodo, etc). [CORGAetq: ANT2005-03-03/27]

1.4. Datos

Da combinación das categorías gramaticais delimitadas no sistema de XIADA cos atributos e valores recoñecidos para cada unha delas derívanse, do noso etiquetario, un total de 453 etiquetas posibles.

Destas 453, 70 só se observan na consulta directa ao lexicón ou na primeira versión do léxico xerado a partir do lexicón⁵⁹, pois nestas etiquetas mantense a ambigüidade nalgún atributo, como por exemplo o valor *determinante/non determinante*. Esta indeterminación na atribución de certos valores é imposible defendela na anotación de corpus, posto que ao estar as formas inmersas nun contexto específico, este determinará a súa caracterización como *determinante* ou *non determinante*, por exemplo, mais nunca poderán estar presentes simultaneamente os dous valores na mesma unidade léxica. É dicir, para facilitar a introdución no lexicón de certos elementos, caracterizamos, por exemplo, o demostrativo *estas* como Eafp (*Demostrativo, determinante/non determinante, feminino, plural*); non obstante, á hora de etiquetar e desambiguar corpus⁶⁰, esta forma demostrativa só pode marcarse como

⁵⁹ Aparecen, por exemplo, no léxico descargable de Xiada na súa versión 2.2 (723.998 entradas) dispoñible en <http://corpus.cirp.es/xiada/descargas.html> [consultado o 21/05/2010]. Esas caracterizacións son incorrectas e figuran porque non detectamos a tempo o fallo do scrip de filtrado que se lle aplicou.

⁶⁰ Lembremos que o noso obxectivo é asignarlle a cada unidade léxica a etiqueta e lema que lle corresponde en cada ocorrencia textual e evitar, sempre que sexa posible, a presenza de ambigüidades nos documentos etiquetados.

Edfp (*Demostrativo, determinante, feminino, plural*) ou Enfp (*Demostrativo, non determinante, feminino, plural*) xa que ambos os valores xuntos son incompatibles.

Na anotación morfosintáctica do *Corpus de Referencia do Galego Actual etiquetado*, versión 2.1, que consta de 386.498 formas ortográficas ou, se se prefire, 426.264 elementos gramaticais, maniféstase unicamente o 78,07% do total, ou sexa 299 etiquetas. Isto non significa que as 84 restantes non sexan útiles ou non poidan aparecer en textos reais, senón que, de momento, debido ás características dos documentos que constitúen este corpus (textos xornalísticos, polo tanto maioritariamente crónicas e predominio das terceiras persoas verbais e persoais) aínda non afloraron. Están recollidas, iso si, no núcleo ou *kernel* de XIADA, o cal debe conter cada etiqueta polo menos unha vez.

No apéndice A, recollemos o etiquetario completo de XIADA e nos anexos B, C e D poden verse, respectivamente, as listaxes completas de etiquetas cuxa aparición se documenta no CORGAetq, as que a maiores se rexistran no núcleo ou *kernel* de XIADA e as que só se empregan na xestión interna do lexicón. Os inventarios, ademais das etiquetas, inclúen un ou varios exemplos.

Finalmente, na táboa que segue resumimos en cifras a información relativa ás distintas etiquetas que é posible atopar no sistema de XIADA, entendendo por este todos os seus compoñentes: lexicón, *kernel* e corpus de adestramento⁶¹. Para que as diferenzas sexan máis facilmente recoñecibles, marcamos con negriña o número de etiquetas que, por categoría gramatical, se recollen no *kernel* e que difire do número de etiquetas presentes no CORGAetq. En total, 453 etiquetas dan conta, en teoría, da caracterización morfolóxica de calquera palabra pertencente á lingua galega e, na práctica, 383 permiten a adecuación desa caracterización a un contexto dado:

⁶¹ O corpus de adestramento do Etiquetador/lematizador do galego actual [<http://corpus.cirp.es/xiada/etiquetador.html>] está dispoñible para a súa consulta pública co nome de *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq) [<http://corpus.cirp.es/corgaetq>]. A idea inicial era incluír no CORGAetq tanto os documentos etiquetados manualmente como os procesados só automaticamente, de xeito que o CORGAetq se convertese nunha especie de corpus parello ó CORGA, coa única diferenza de que un estaría anotado e o outro non. Non obstante, recapacítamos ao pensar no distinto grado de fiabilidade da anotación, a cal esixía manter sistemas distintos. Deste xeito, cando se considere que o etiquetador posúe suficiente corpus para o seu adestramento, este pecharase e, con el, o CORGAetq; pola súa banda, o *Corpus de Referencia do Galego Actual etiquetado automaticamente* (CORGAetqa), que estará accesible en <http://corpus.cirp.es/corgaetqa>, acollerá os textos etiquetados só automaticamente, sen revisión manual, e irase incrementando paulatinamente cos documentos dos distintos medios que conforman o CORGA, a medida que se vaia rematando o adestramento de cada módulo.

Categoría	CORGAetq	XIADA- kernel	XIADA- lexicón
Adverbio	4	4	1
Adxectivo	16	21	
Artigo	8	8	
Categoría periférica	12	18	
Conxunción	2	2	
Demostrativo	9	9	4
Indefinido	17	18	3
Interrogativo-exclamativo	11	14	7
Interxección	1	1	
Locución	4	4	
Numeral	18	18	9
Posesivo	38	56	26
Preposición	1	1	
Pronome	60	87	13
Relativo	11	11	4
Signo de puntuación	18	20	
Substantivo	14	16	
Verbo	55	75	3
Total	299	383	70

1.5. Outros etiquetarios para o galego

Cando se determinou a primeira versión do etiquetario de XIADA (Vilares et al., 1998) non existía para o galego ningún outro. Hoxe en día, á versión definitiva do *tag-set* de XIADA, aplicado na etiquetación do *Corpus de Referencia do Galego Actual etiquetado*, podemos engadir tamén os dous etiquetarios deseñados polo Seminario de Lingüística Informática (SLI) da Universidade de Vigo: un deles (Aguirre Moreno et al. 2001, 2002a e 2002b) aplicado ao corpus de traducións TECTRA; o outro (Gómez Guinovart e López Fernández, 2009) aplicado ao *Corpus Técnico Anotado do Galego (CTAG)*⁶².

⁶² Non incorporamos á comparación realizada o empregado no *Tesouro Informatizado da Lingua Galega* (TILG), tanto na versión dispoñible en <http://www.ti.usc.es/tilg/> como na de <http://sli.uvigo.es/TILG/index.html>, por dous motivos: 1) non existe ningún traballo que nós coñezamos que explique o conxunto de etiquetas empregado; 2) non delimita clases de palabras nin atributos funcionais para cada unha delas, senón que se limita a listar as etiquetas completas e os códigos que corresponden a cada unha. Por exemplo, para o Artigo, entendemos, as "Cat_Letra" que aparecen son as seguintes: ArtD_ms, ArtD_ms*, ArtD_fs, ArtD_fs*, ArtD_mp, ArtD_mp*, ArtD_fp, ArtDfp*, ArtI_ms, ArtI_ms*, ArtI_fs, ArtI_fs*, ArtI_mp, ArtI_mp*, ArtI_fp, ArtI_fp*. Agora ben, co uso da maiúscula debemos deducir que se consideran categorías diferenciadas o Artigo determinado ou definido e o Artigo indeterminado ou indefinido? Por que o asterisco aparece na etiqueta se en realidade o que se quere indicar é que as formas con ela etiquetadas son variantes? Todo o que poderíamos facer serían conxecturas e non unha comparación.

Creemos que pode ser interesante realizar unha comparación entre o etiquetario de XIADA que acabamos de describir e os dous do Seminario de Lingüística Informática de Vigo, posto que todos eles seguen as recomendacións de EAGLES e están funcionando na actualidade para etiquetar corpus lingüísticos de galego.

O etiquetario de XIADA e o aplicado ao corpus TECTRA, que de aquí en diante denominaremos SLI-TECTRA, fundaméntanse principalmente na descrición gramatical de Álvarez, Regueira e Monteagudo (1986), mentres que para o empregado no CTAG adaptáronse as propostas realizadas por Civit (2003) para o castelán.

Na táboa que segue, partindo da proposta de EAGLES recollida en Monachini e Calzolari (1996), reflectimos as categorías delimitadas en cada un dos etiquetarios do galego que imos contrastar: SLI-TECTRA, SLI-CTAG e XIADA. Para a listaxe das clases de palabras respectamos a orde establecida en EAGLES e se nalgún dos etiquetarios non existe esa clase deixamos o espazo en branco.

	EAGLES (14)	SLI (TECTRA) (13)	SLI (CTAG) (21)	XIADA (18)
1.	Nome	Nome	Nome	Substantivo
2.	Adxectivo	Adxectivo	Adxectivo	Adxectivo
3.	Pronome	Pronome	Pronome	Pronome
4.	Verbo	Verbo	Verbo	Verbo
5.	Artigo	Artigo	Determinante	Artigo
6.	Interxección	Interxección	Interxección	Interxección
7.	Conxunción	Conxunción	Conxunción	Conxunción
8.	Adverbio	Adverbio	Adverbio	Adverbio
9.	Adposición	Preposición	Preposición	Preposición
10.	Puntuación	Puntuación	Puntuación	Puntuación
11.	Residual	Residual		Periférica
12.	Numeral		Numeral	Numeral
13.	Única	Partícula negativa		
14.	Determinante			
15.				Locución
16.			Demostrativo	Demostrativo
17.			Posesivo	Posesivo
18.			Indefinido	Indefinido
19.			Relativo	Relativo
20.			Interrogativo	Interrogativo-exclamativo
21.			Cifra	
22.			Abreviatura	
23.			Símbolo	
24.			Estranxeirismo	
25.			Palabra non clasificada	

Observando a táboa, o primeiro que percibimos é o distinto número de clases de palabras de cada un dos etiquetarios e a semellanza categorial por pares: EAGLES/SLI-TECTRA fronte a SLI-CTAG/XIADA.

Se prescindimos das distintas denominacións para as categorías 9 (Adposición vs. Preposición) e 13 (Única vs. Partícula negativa), as categorías consideradas no SLI-TECTRA son idénticas ás propostas por EAGLES, coa salvidade de que non reflicten no seu etiquetario a distinción establecida en Monachini e Calzolari (1996) entre as clases Pronome e Determinante. No SLI-TECTRA optaron por non empregar a distinción entre pronome e determinante, seguindo as descrições gramaticais do galego nas que se basearon para a elaboración do etiquetario; isto é, seguindo a Álvarez, Regueira e Monteagudo (1986) en Aguirre Moreno et al. 2002a (pp. 24, 28) e a Álvarez, Regueira e Monteagudo (1986) e Álvarez e Xove (2002) en Aguirre Moreno et al. 2002b (pp. 210, 217).

A única categoría que non presenta diferenzas nos tres etiquetarios é a Interxección, que carece de atributos nos tres.

Na clase Preposición a diverxencia procede do sistema de XIADA por remitir as preposicións multipalabra ao subtipo *preposicional* da categoría Locución.

No que respecta á Conxunción, se non temos en conta que en XIADA as conxuncións multipalabra se caracterizan como Locución conxuntiva, coinciden o SLI-CTAG e XIADA, que diferencian no tipo entre coordinante e subordinante, e discrepa o SLI-TECTRA que non lle reconece ningún atributo:

	SLI-TECTRA	SLI-CTAG	XIADA
Categoría	Conxunción	Conxunción	Conxunción
Atributos	Sen atributos	coordinativa subordinativa	coordinante subordinante

Curiosamente, tamén nos signos de puntuación delimitados hai diferenzas nos tres sistemas. O SLI-TECTRA non recolle o signo dobre de tipo parentético *chaves* ({}), e o SLI-CTAG, seguindo o etiquetario utilizado no Freeling, que á súa vez se basea en Civit⁶³, inclúe a maiores o símbolo da porcentaxe (%) e os que para nós son símbolos matemáticos claros: os operadores +, -, =. Na táboa que segue establecemos os

⁶³ En Civit (2003: 122-123) si está caracterizado como signo de puntuación % pero non aparecen por ningures os símbolos da suma, resta e igualdade.

elementos discordantes a partir dos signos ortográficos recollidos no apartado 1.3.2.17. deste capítulo:

	SLI-TECTRA	SLI-CTAG	XIADA
Categoría	Puntuación	Puntuación	Puntuación
Atributos	Sen atributos	Sen atributos	Sen atributos
Elementos discordantes	{ }	% - + =	

O SLI-TECTRA é o único que individualiza o adverbio *non* ao clasificalo como Partícula negativa. Tanto en XIADA como no SLI-CTAG se trata na clase Adverbio.

A categoría Cifra só a reconece o SLI-CTAG e remite a ela anos, enderezos, números de teléfono, etc. En XIADA xa vimos (epígrafe 1.3.2.7. Numeral) os motivos que nos levaron a eliminar a categoría Cifra e como recategorizamos os elementos que a ela se remitían. No etiquetario do SLI-TECTRA (Aguirre Moreno et al., 2001, 2002a e 2002b) só podemos supoñer que se analizan nos numerais, pois a) non hai mención ningunha a unha categoría ou valor Cifra; b) non se describen as distintas categorías máis que a través de táboas nas que se recollen os atributos e valores que son pertinentes xunto con algún exemplo e c) non atopamos ningún exemplo cunha cifra nin nas táboas nin no exemplo reproducido dunha unidade etiquetada morfosintacticamente.

A categoría Residual de EAGLES e do SLI-TECTRA, Periférica en XIADA, desaparece como tal no SLI-CTAG. No seu lugar preferiron considerar novas clases de palabras algúns dos subtipos recollidos en EAGLES; isto é, no SLI-CTAG, a clase Residual desagregase nas categorías Abreviatura, Símbolo, Estranxeirismo e Palabra non clasificada. En ningún dos dous etiquetarios do SLI –dá igual que se trate de distintas categorías (SLI-CTAG) ca de valores do atributo *tipo* (SLI-TECTRA)–, en contra do que sucede no etiquetario de XIADA, se teñen en conta os atributos de xénero ou número, pertinentes, na nosa opinión, para caracterizar morfoloxicamente, por exemplo, abreviaturas e estranxeirismos.

Na táboa seguinte mostramos as adaptacións que realizaron os tres sistemas partindo da mesma base: a categoría Residual de EAGLES.

Etiquetario	EAGLES			SLI-TECTRA		SLI-CTAG			XIADA		
Categoría	Residual			Residual		Abreviatura			Periférica		
						Símbolo					
						Estranxeirismo					
						Palabra non clasificada					
Atributos	Tipo	x ^o	n ^o	Tipo				Tipo	x ^o	n ^o	
Valores	foreign word	m	sg	palabra estranxeira				fórmula	m	sg	
	formula	f	pl	fórmula				símbolo	f	pl	
	symbol	n	n	símbolo				sigla	a (m/f)	a (sg/pl)	
	acronym	c		acrónimo				abreviatura	0	0	
	abbreviation			abreviatura				outros tipos			
	unclassified			sen clasificar							

Nas varias publicacións onde se dá a coñecer o etiquetario do SLI-TECTRA (2001, 2002a, 2002b) non se describen en ningún momento os criterios de delimitación das distintas categorías recoñecidas, nin como se aplican os atributos, nin se proporciona información detallada sobre os elementos que conforman os diferentes subtipos na categoría Residual. A falta de detalles e fixándonos nos subtipos e exemplos, intuimos, *grosso modo*, cales son os criterios empregados e a que elementos poden caracterizar.

No SLI-CTAG inclúese unha descrición mínima para cada clase gramatical, mais nela si se dá conta das características esenciais da clase e de que elementos a integran. Para a clase Abreviaturas aclárase:

No CTAG emprégase a etiqueta Y para as abreviaturas de resolución incerta e tamén para os enderezos electrónicos e indicacións de unidades de temperatura (°C) e outras. Porén, etiquétanse como nomes propios formas como *M^a* ou siglas que corresponden a entidades propias e individualizadas, como *CEE* ou *EEUU*; como nomes comúns formas abreviadas como *n^o* ou *ex.* ou siglas do tipo *SA*, *SP* ou *PEMES* (sic); e como numerais ordinais as abreviaturas como *1^o* ou *3^o*. [Gómez Guinovart e López Fernández, 2009: 67]

Nas categorías do SLI-CTAG paralelas á Residual de EAGLES chámannos a atención varios feitos:

1) A non aparición da clase Sigla ou Acrónimo. Por moito que, gramaticalmente, as siglas poidan ser tratadas como substantivos, ben comúns ben propios, o non recoñecelas implica que non poden ser individualizadas e, polo tanto, alguén interesado en realizar un estudo sobre elas, extraendo o material lingüístico do CTAG, atoparía o labor bastante tedioso pois requiriría unha consulta previa sobre todos os substantivo para, a continuación, peneirar manualmente os resultados e quedar só coas ocorrencias das siglas.

2) A remisión á clase Símbolo só dos elementos químicos da táboa periódica ou formas compostas por estes. Isto obrígaos a encaixar noutras categorías elementos que todos consideramos símbolos, como son os operadores matemáticos (trátanos nos signos de Puntuación) ou unidades de temperatura e “outras” (inclúenas nas Abreviaturas).

3) A incoherencia presente na delimitación da clase Abreviatura por caracterizar como substantivos comúns abreviaturas frecuentes coma *n^o* ou *ex.* e, non obstante, caracterizar como abreviaturas elementos que non se abrevian (enderezos electrónicos) ou unidades de medida recoñecidas internacionalmente como símbolos.

Para o Artigo, a principal diverxencia procede do SLI-CTAG no que, deixando de lado a denominación de Determinante fronte á de Artigo no SLI-TECTRA ou en XIADA, que non reviste maior importancia, só inclúe nesta clase o artigo determinado, pois o indeterminado remítene á categoría Indefinido. A táboa que segue reflicte as caracterizacións das formas do artigo nos tres etiquetarios:

Etiquetario	SLI-TECTRA	SLI-CTAG		XIADA
Categoría	Artigo	Determinante	Indefinido	Artigo
Tipo	definido indefinido			determinado indeterminado
Xénero	masculino feminino	masculino feminino común	masculino feminino	masculino feminino
Número	singular plural	singular plural	singular plural	singular plural
Formas	o, a, os, as, un, unha, uns, unhas	o, a, os, as, @s	un, unha, uns, unhas	o, a, os, as, lo, la, los, las, un, unha, uns, unhas

Como pode observarse, o de XIADA é o único que individualiza as denominadas segundas formas do artigo co que, en caso de interesarlle a alguén, pódese

estudar, por exemplo, a frecuencia e casuística do emprego das segundas formas antes e despois dos cambios introducidos no 2003 polas *Normas* (RAG/ILG, 2003).

En cambio, o SLI-CTAG é o único que reconece como forma do artigo @s, caracterizado morfoloxicamente como GCP (*Determinante, común, plural*). A presenza nesta forma da arroba, antiga medida de capacidade e masa, que hoxe en día se utiliza para referirse conxuntamente a individuos de ambos os sexos “*para evitar las engorrosas repeticiones a que da lugar la reciente e innecesaria costumbre de hacer explícita la alusión a los dos sexos*” (DPD, 2005: 311) parécenos unha solución, cando menos, incompleta.

Nós non tivemos que enfrontarnos aínda coa aparición do símbolo arroba como substituto das vogais, “a”, “e” e “o”, motivo polo que non dispoñemos de exemplos do CORGAetq. Non obstante, cremos que a proposta de resolución debe ir por outro camiño. En primeiro lugar, a arroba é un símbolo, non un signo lingüístico; en segundo lugar, o emprego da arroba non se constata só no artigo determinado senón tamén en substantivos, adxectivos, indefinidos, etc., como demostran os seguintes exemplos extraídos do CORGA:

Para tod@s el@s vai este artigo. [CORGA: ANT2002-11-07]

O longo debate comeza por saber cal será o sexo do terceir@ en cuestión. [CORGA: C. Franco; *Palabras contadas*, Xerais, 2006]

Mais, ao crearmos formas novas (como “médica” ou “xuíza”, usando a arroba para dicir “@s mestr@s”, “@s nen@s” ou preferindo falar dos “dereitos humanos” e non dos “dereitos do home”) tamén nos rebelamos, e de forma ben evidente, a facer invisible á metade da humanidade. [CORGA: T. Moure; *Outro idioma é posible. Na procura dunha lingua para a humanidade*, Galaxia, 2005]

Xunt@s temos futuro. [CORGA: CC2006-01]

A solución que se adopte, en consecuencia, debe ser aplicada ás distintas categorías e non só ao artigo. Que decisión tomar entón? Cremos que a solución está en non caracterizar categorialmente estas formas con arroba, senón en que, de xeito semellante ás contraccións, se deben desagregar nos seus constituíntes, como facemos cando a unha forma masculina se lle engade unha barra e logo a terminación feminina. Por exemplo, no CORGAetq:

Levo ben anos en contacto con alumnos/as, con ensinantes, bibliotecarias ou promotores de lectura tentando convencelos dos beneficios indiscutibles da literatura. [CORGAetq: ANT2008-07-10/25]

Análise unidade ▾
Unidade
alumnos/as
Constituínte ▾
Forma alumnos Etiqueta Scmp Lema alumno
Constituínte ▾
Forma / Etiqueta Q/ Lema /
Constituínte ▾
Forma alumnas Etiqueta Scfp Lema alumno

Deste xeito, non cometemos a incongruencia de incorporar o símbolo arroba ás grafías do alfabeto galego; non nos vemos na obriga de introducir e caracterizar tres unidades por cada substantivo, adxectivo, etc. que manifeste oposición formal de xénero (*nen*, *nena*, **nen@*) e facilitamos o estudo ou consulta do recurso que se creou tamén en galego para liberar a lingua de estereotipos discriminatorios.

Na clase Adverbio o etiquetario de XIADA desmárcase totalmente dos deseñados polo SLI-TECTRA e o SLI-CTAG, pois nela caracterizamos como adverbios, ben relativos ben interrogativo-exclamativos, as formas *como*, *cando* e *onde* e, nos demais elementos do paradigma, indicamos se son nucleares ou actúan de modificadores. Diferenciamos, ademais, entre unidades simples (Adverbio) e unidades multipalabra (Locución adverbial). E, finalmente, a nivel de etiqueta non se individualizan os adverbios con sufixo apreciativo nin se marca a gradación. Estes fenómenos poden localizarse no CORGAetq coa combinación de *forma*, *etiqueta* e *lema*, como se pode comprobar no seguinte exemplo:

Non é lonxísimo, está cerquiña: [lonxísimo Wn lonxe], [cerquiña Wn cerca]

No etiquetario do SLI-CTAG o único atributo que caracteriza o adverbio, sexa este simple ou multipalabra, é o *grao*; mais con el non se refíren á gradación do adverbio a semellanza da do adxectivo, senón ao emprego do sufixo intensificador *-iño*; polo tanto, cando aparece o apreciativo, o adverbio caracterízase como RA, mentres que no resto dos casos o valor é *non especificado*, R0. Curiosamente, dos dous exemplos que achegan para o uso do apreciativo, *loguiño*, *a_modiño*, o primeiro remítese ao lema *logo*, en tanto que o segundo mantén a mesma expresión na unidade e no lema,

a_modiño (Gómez Guinovart e López Fernández, 2009: 64). En relación co atributo *grao*, chámanos así mesmo a atención o feito de que morfoloxicamente dean conta do adverbio apreciativo e que non o fagan igualmente coas ocorrencias do superlativo, xa que ningún dos dous sufixos é posible empregalo con todos os adverbios e tan frecuente é un coma o outro.

Polo que respecta ao etiquetario do SLI-TECTRA debemos diferenciar entre a versión do 2002a, publicada na revista *Procesamiento del Lenguaje Natural* (PLN), onde a categoría Adverbio non presenta atributos, da de 2002b, publicada en *Viceversa*, onde aplican os atributos *grao* (cos valores *positivo*, *comparativo* e *superlativo*) e *forma* (valores *plena* e *diminutivo*, empregándose este último só nos adverbios de *grao positivo*).

En ningún dos dous etiquetarios do SLI sabemos, non obstante, como tratan as formas *como*, *cando* e *onde*, pois non aparecen recollidas nin exemplificadas en ningunha das clases ás que poderían remitirse: Adverbios, Interrogativos, Relativos.

En resumo, a clasificación dos adverbios nos distintos etiquetarios responde aos seguintes atributos:

Etiquetario	SLI-TECTRA 2002a	SLI-TECTRA 2002b		SLI-CTAG	XIADA
Categoría	Adverbio	Adverbio		Adverbio	Adverbio
Atributos		Grao	Forma	Grao	Tipo
Valores		positivo	plena diminutivo	apreciativo	nuclear
		comparativo	plena	non especificado	modificador
		superlativo	plena		nuclear/modificador
					relativo
					interrogativo-exclamativo

Como veremos a continuación, onde máis diverxencias se dan entre os tres etiquetarios que estamos contrastando é nas clases que EAGLES denomina Pronome e Determinante.

A clasificación categorial das clases de EAGLES Pronome e Determinante levada a cabo polo SLI-CTAG, deixando de lado de momento os atributos, coincide coa

de XIADA, salvo na inclusión do artigo indefinido, como xa vimos, na clase Indefinido. Gómez Guinovart e López Fernández (2009: 64) indican na clase Pronome que “*Malia que Civit (2003) inclúe nesta categoría os pronomes demostrativos, posesivos, indefinidos, relativos, interrogativos, exclamativos e numerais, na etiquetaxe do CTAG todas estas categorías considéranse categorías independentes, reservándose a categoría pronominal do etiquetario para os denominados tradicionalmente pronomes persoais*”. Pois ben, o que eles non sinalan é que Civit⁶⁴ (2001: 23-24; 2003: 74-103) realiza unha dobre clasificación dos demostrativos, posesivos, indefinidos, interrogativos, exclamativos e numerais delimitándoos na clase dos Determinantes e na dos Pronomes dependendo de se funcionan como determinantes ou como núcleos, respectivamente. Así pois, o SLI-CTAG compórtase igual ca o SLI-TECTRA e non reconece a dobre funcionalidade dos tradicionalmente denominados determinativos, só que neste último prefírese consideralos subtipos do Pronome (Pronome demostrativo, Pronome posesivo, etc.) e non categorías independentes.

É XIADA, pois, o único etiquetario galego no que co atributo *valor* (*determinante, non determinante*) se respecta a recomendación de EAGLES de diferenciar entre valores determinantes e pronominais nos demostrativos, posesivos, indefinidos, relativos, interrogativos-exclamativos e numerais.

De seguido veremos os restantes atributos que cada un dos etiquetarios aplica para cada unha das categorías⁶⁵.

No Demostrativo, ademais dos atributos *xénero* (*masculino, feminino, neutro*) e *número* (*singular, plural*) presentes nos tres etiquetarios, o SLI-TECTRA (2002a, 2002b) engade a *persoa* (1, 2, 3). Así, por exemplo, a forma demostrativa *ese* recibe, dependendo do etiquetario, as seguintes etiquetas:

SLI-TECTRA: PDMS2 (*Pronome demostrativo masculino singular segunda*)

SLI-CTAG⁶⁶: DMS [*Demostrativo, masculino, singular*]

⁶⁴ Lembremos que para a determinación do etiquetario se basean en Civit (2003), quen á súa vez segue a proposta de EAGLES.

⁶⁵ Excepto no pronome persoal, imos denominar categoría o que para SLI-TECTRA constitúe un subtipo do Pronome co obxecto de non repetir constantemente aquilo de “categoría para SLI-CTAG e Xiada, subtipo no SLI-TECTRA”.

⁶⁶ Empregamos os corchetes para sinalar que o desenvolvemento da etiqueta no SLI-CTAG é sempre obra nosa, xa que no artigo onde se describe o etiquetario empregado no CTAG non figura ningunha etiqueta desenvolva.

XIADA: Edms (*Demostrativo, determinante, masculino, singular*)

XIADA: Enms (*Demostrativo, non determinante, masculino, singular*)

Para o Posesivo, excepto no SLI-TECTRA onde incorporan tamén o atributo *cortesía* (cortés), os demais atributos (persoa, xénero, número, e número do posuidor) reflíctense nos tres etiquetarios, só que a súa orde, valores ou aplicación varían dun a outro.

Na táboa que segue recollemos as tres caracterizacións, indicando nas celas os valores dos atributos e a posición que ocupan na cadea de etiquetaxe de cada un dos sistemas.

Etiquetario	SLI-TECTRA	SLI-CTAG	XIADA
Categoría	Pronome 1	Posesivo 1	Posesivo 1
Tipo pronome	posesivo 2		
Valor			determinante non determinante determinante/non determinante 2
Persoa	1 ^a , 2 ^a , 3 ^a 4	1 ^a , 2 ^a , 3 ^a 2	1 ^a , 2 ^a , 3 ^a 3
Xénero	masculino feminino 6	masculino feminino 3	masculino feminino 5
Número	singular plural 7	singular plural 4	singular plural 6
Número do posuidor	singular plural 3	singular plural 5	singular plural singular/plural 4
Cortesía	cortés 5		

No SLI-TECTRA a lonxitude da cadea de etiquetaxe no posesivo varía dos 6 aos 7 caracteres, pois só cando a forma é *cortés* cobren o oco correspondente; se non o é, non se indica de ningunha maneira na etiqueta nin no desenvolvemento desta. Confróntense, por exemplo, as etiquetas pertencentes a *seu* referido a *el/ela* e a *vostede*:

seu PVS3MS (*pronome posesivo. Posuidor(a): singular terceira. Posuído(a): masculino singular*)

seu PVS2CMS (*pronome posesivo. Posuidor(a): singular segunda de cortesía. Posuído(a): masculino singular*)

No SLI-CTAG cando o posuidor é unha terceira persoa (*seu*) asígnanlle como valor ao número do posuidor 0, “*dada a dificultade de distinguir o seu número gramatical singular ou plural, isto é, se fai referencia a el/ela ou a eles/elas*” (Gómez Guinovart e López Fernández, 2009: 65).

No SLI-TECTRA debemos acudir a Aguirre Moreno et al. (2001) para entender por que no número do posuidor, aplicado á terceira persoa, non existe un valor ambiguo:

En cuanto a la subespecificación léxica de género o número, es decir, aquellas palabras que formalmente no presentan oposición de género (*cantante*) o de número (*martes*), no reciben ninguna etiqueta que las distinga de las demás. Este criterio es aplicable a todas las categorías. Simplemente se les asigna el género o número que presenten en cada contexto. En este sentido nuestro etiquetario tiene una clara orientación a la anotación de corpus, frente a la marcación de información léxica. [Aguirre Moreno et al., 2001: 15]

A explicación é clara, asígnaselle o que presente en cada contexto. Con todo, seguimos atopando un problema na asignación do número do posuidor ás formas da terceira persoa, porque ou ben o etiquetador é moi potente e ten en conta un contexto moi amplo, ou ben, se está limitado por exemplo á oración, a ambigüidade existe: un lingüista non sempre é capaz de deducir o número do posuidor e xa non digamos un etiquetador automático.

Ante unha oración coma “*O seu volume de negocio aumentou arredor do 15%*”, na que non sabemos se *seu* fai referencia a un singular ou a un plural, ignoramos que criterio se segue no SLI-TECTRA para escoller entre as dúas posibles etiquetas; si sabemos, non obstante, cales lle corresponderían no SLI-CTAG e en XIADA:

SLI-TECTRA: PVS3MS (*pronome posesivo. Posuidor(a): singular terceira. Posuído(a): masculino singular*)

SLI-TECTRA: PVP3MS (*pronome posesivo. Posuidor(a): plural terceira. Posuído(a): masculino singular*)

SLI-CTAG: XM3S0 [*Posesivo, masculino, terceira, singular, non especificado*]

XIADA: Md3ams (*Posesivo, determinante, 3ª, singular/plural, masculino, singular*)

Nas clases Interrogativo e Relativo, á parte do atributo *valor* presente en XIADA para indicar cal é a función dos elementos tradicionalmente denominados

determinativos, os atributos restantes, *xénero* e *número*, coinciden nos tres etiquetarios, malia que os valores e caracterización destes difire nos tres.

O SLI-TECTRA, igual que sucedía no Posesivo co número do posuidor para as formas de terceira, só inclúe os valores *masculino* e *feminino* para o xénero e *singular* e *plural* para o número, malia que logo aparecen etiquetas e exemplos sen que presenten ningún valor para xénero ou número:

quen	PR	<i>pronome relativo (invariable)</i>
	PIN	<i>pronome interrogativo (invariable)</i>
cal (o)	PRMP	<i>pronome relativo masculino singular</i>
	PINMS	<i>pronome interrogativo masculino singular</i>

O SLI-CTAG, ademais de ser o único que introduce o atributo *grao* co valor *apreciativo* nas formas interrogativas e relativas, incorpora o valor *común* para o xénero e *invariable* para o número, a semellanza do que se fai en XIADA cos valores ambiguos *masculino/feminino* e *singular/plural*. Porén, a diferenza estriba na súa aplicación: mentres que no SLI-CTAG a unidade *que*, por exemplo, sempre se etiqueta como TCN0 [*Interrogativo, común, invariable, non especificado*] se é interrogativo, ou QCN0 [*Relativo, común, invariable, non especificado*] se é relativo, en XIADA especificase o xénero e número sempre que o contexto o permite:

Que día tan bo! [Gdms] (*Interrogativo-exclamativo, determinante, masculino, singular*)

Que camisa máis bonita! [Gdfs] (*Interrogativo-exclamativo, determinante, feminino, singular*)

Que é? [Gnas] (*Interrogativo-exclamativo, non determinante, masculino/feminino, singular*)

Na categoría Indefinido podemos afirmar que se repite a situación anterior.

O SLI-TECTRA⁶⁷ inclúe nas táboas definatorias os valores *masculino* e *feminino* para o xénero; *singular* e *plural* para o número; *positivo* e *superlativo* para o grao e *plena* e *diminutivo* para o atributo *forma*, malia que logo aparecen etiquetas e exemplos sen que presenten ningún valor para eses atributos:

algo	PI	<i>pronome indefinido (invariable)</i>
algún	PIMS	<i>pronome indefinido masculino singular</i>

⁶⁷ A versión de 2002b incorpora, a maiores da de 2002a, os atributos *grao* (*positivo, superlativo*) e *forma* (*plena, diminutivo*) para caracterizar os indefinidos.

pouquiño PIDFS *pronomes indefinidos diminutivos masculino singular*

O SLI-CTAG, categorialmente, só reconece os apreciativos, non os indefinidos superlativos, pois para o atributo *grao* unicamente figura o valor *apreciativo*. Consultamos o CTAG⁶⁸ para investigar que tratamento reciben formas coma *mesmísimo*, *moitísima*, etc. e percibimos que o lema dos superlativos é a forma masculina singular: *moitísimo*, *mesmísimo*, etc. Polo que respecta aos valores do xénero e do número, sorprende a aparición do *neutro*, aplicado á caracterización dos indefinidos nucleares invariables *calquera* e *ninguén*. Descoñecemos que criterio seguen para o seu emprego pois, curiosamente, a *alguén*, *algo* e *nada* asígnaselles o xénero masculino (Gómez Guinovart e López Fernández; 2009: 66):

algo	IMS0	[Indefinido, masculino, singular, non especificado]
alguén	IMS0	[Indefinido, masculino, singular, non especificado]
ninguén	INS0	[Indefinido, neutro, singular, non especificado]
pouquiño	IMSA	[Indefinido, masculino, singular, apreciativo]

En XIADA, unha vez máis, percíbese a intención de proporcionar unha análise morfosintáctica detallada, tanto na descrición das formas no lexicón como na súa etiquetación en corpus. Así, ademais do atributo *valor* que nos permite saber se unha forma funciona como determinante ou como nuclear, tamén se manteñen nos atributos *xénero* e *número* os valores individuais e os ambiguos para tratar de caracterizar cada forma invariable segundo as concordancias que se establezan no contexto. Deste xeito, empregamos os valores *masculino* e/ou *feminino* e *singular* na caracterización morfosintáctica dos indefinidos referidos a persoa que sempre son nucleares: *alguén*, *ninguén*, *cadaquén* e *quenquera*; pola contra, nas formas unicamente nucleares que posúen o trazo non humano (*algo*, *nada*, *ren* e *res*) empregamos unha etiquetación dobre: unha, cos valores *masculino singular* (Inms), destinada a caracterizar os casos nos que se explicita esta concordancia e outra, onde non aplica o xénero nin o número (In00), reservada para aqueloutros nos que non se observe ningún tipo de concordancia:

Advirten de que todo é Made in USA, nada importado. [Inms] (*Indefinido, non determinante, masculino, singular*)

⁶⁸ No CTAG só se pode consultar por forma ou lema, combinándoos coa categoría gramatical e, nalgúns casos, co patrón gramatical. En ningún caso podemos ver a etiqueta que caracteriza morfoloxicamente a unidade de busca.

El e mais eu non nos parecemos en nada. [In00] (*Indefinido, non determinante, xénero non aplica, número non aplica*)

Polo que se refire á identificación dos indefinidos cando aparecen co sufixo intensificador (-*iño*) ou superlativo (-*ísimo*), en XIADA, como xa vimos na comparación da categoría Adverbio, a nivel de etiqueta non se individualizan categorialmente, pero si poden localizarse no CORGAetq coa combinación de *forma*, *etiqueta* e *lema*, como se pode comprobar no seguinte exemplo:

Pedinche unha pouquiña auga, non moitísima:

pouquiña: [Idfs] (*Indefinido, determinante, feminino, singular, (lema) pouco*)

moitísima: [Infs] (*Indefinido, non determinante, feminino, singular, (lema) moito*)

Finalmente, para rematar coas categorías implicadas nas clases Pronome e Determinante de EAGLES, quédanos contrastar o tratamento que recibe o pronome persoal nos etiquetarios do SLI e de XIADA. A seguinte táboa recolle os atributos e valores empregados en cada un dos sistemas de anotación:

Etiquetario	SLI-TECTRA	SLI-CTAG	XIADA
Categoría	Pronome 1	Pronome 1	Pronome 1
Tipo	persoal 2		
Tonicidade	tónico, átono 7		tónico, átono 2
Persoa	1 ^a , 2 ^a , 3 ^a 5	1 ^a , 2 ^a , 3 ^a 2	1 ^a , 2 ^a , 3 ^a 4
Xénero	masculino, feminino, neuro 3	masculino, feminino, común 3	masculino, feminino, masculino/feminino 5
Número	singular, plural 4	singular, plural, invariable 4	singular, plural, singular/plural 6
Caso	nominativo, acusativo, dativo, oblicuo, nominativo/oblicuo, dativo/acusativo 8	nominativo, acusativo, dativo, oblicuo, reflexivo, nominativo/recto, acusativo/dativo/reflexivo 5	nominativo, acusativo, dativo, preposicional, formante léxico, nominativo/preposicional ⁶⁹ , outros casos 3
Cortesía	familiar, cortés 6	cortés, non especificado 6	
Tipo de pronome	persoal, persoal exclusivo, reflexivo 6		

⁶⁹ Este valor só se emprega en Xiada para caracterizar algúns pronomes no lexicón (*ti*, por exemplo). En corpus a ambigüidade desfai sempre, pois o contexto indica se é *nominativo* ou *preposicional*.

No SLI-TECTRA, unha vez máis, a información proporcionada difire da táboa na que se representa a xerarquía de subcategorías, neste caso do pronome persoal, ás táboas nas que se recolle o inventario de etiquetas, o seu desenvolvemento e exemplos. Así, o valor *familiar* do atributo *cortesía*, caracterizador de *ti*, *che* e *te*, non se emprega nas etiquetas correspondentes nin no seu desenvolvemento; o atributo *xénero* unicamente se aplica naqueles pronomes que presentan formas diferenciadas para o masculino, feminino ou neutro: *el*, *ela*, *elo*, *eles*, *elas*, *nosoutros*, *nosoutras*, *vosoutros*, *vosoutras*, *o*, *a*, *os*, *as*. O valor da *tonicidade* só aparece nos átonos e o *tipo de pronome* só se marca no *exclusivo* e no *reflexivo*. Polo tanto, establecer unha secuenciación coherente dos atributos para o pronome persoal no sistema do SLI-TECTRA resulta imposible, pois a extensión das cadeas de etiquetaxe persoais difire duns elementos a outros e o lugar que debía ocupar o carácter referido ao *xénero*, *tonicidade*, *cortesía* ou *tipo de pronome* aparece ocupado polo seguinte valor que se considera pertinente. A secuenciación que aquí propoñemos está extraída da caracterización dun dos pronomes para os que son pertinentes todos os atributos, a cal imos reflectir porque cremos que é digna de ser recollida⁷⁰:

o PPMS2CAA

1) A lectura que nós realizamos tendo en conta a etiquetación do paradigma pronominal átono é: *pronome, persoal, masculino, singular, segunda de cortesía, átono, acusativo*

2) O desenvolvemento da cadea de etiquetaxe que aparece na táboa dos pronomes persoais átonos, pola contra, é (Aguirre Moreno et al., 2002a: 31; 2002b: 221): *pronome persoal átono masculino singular segunda de cortesía acusativo*

É evidente que o que nos deixa perplexos é a consideración de forma cortés para o pronome átono de terceira acusativo, *o(s)*, *a(s)*, cando, á parte de todo o argumentado en contra da existencia dunha segunda persoa cortés no verbo e no pronome na epígrafe 1.3.2.2. Verbo, na mesma *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986), obra lingüística na que se fundamenta principalmente o etiquetario, se especifica:

Cómpre advertir que, a pesar de tratarse de formas de segunda persoa, *vostede* e *vostedes* levan o verbo en terceira persoa, singular ou plural, e que as formas oblicuas correspondentes,

⁷⁰ Inmediatamente antes figura a análise, esperable, de “o”: PPMS3AA (*pronome persoal átono masculino singular terceira acusativo*)

tónicas ou átonas, son as destoutra persoa: *¿Vostede quere que o leve a algún sitio ou que lle deixe o coche?* [Álvarez, Regueira e Monteagudo, 1986: 161-162]

Á parte do xa dito para o SLI-TECTRA, podemos indicar que nos átonos só se lles asigna *caso* aos pronomes que, pola súa forma, non presentan ambigüidade: *acusativo* [te, o(s), a(s)] e *dativo* [che, lle(s)]. Para *me*, *nos* e *vos* indícase a persoa e o número e para *se*, ademais, que é reflexivo e pode ser *cortés*. No paradigma pronominal tónico, para aquelas formas que se caracterizan como rectas e oblicuas nas gramáticas (*ti*, *el*, *vostede*, etc.) achéganse análises dobres, unha por cada. Ex.:

ti PPS2 pronome persoal tónico recto singular segunda

ti PPS2O pronome persoal tónico oblicuo singular segunda

Co que acabamos de ver, temos que mostrar o noso total desacordo en relación coa afirmación realizada en Aguirre Moreno et al. (2001):

En cuanto a la subespecificación léxica de género o número, es decir, aquellas palabras que formalmente no presentan oposición de género (*cantante*) o de número (*martes*), no reciben ninguna etiqueta que las distinga de las demás. Este criterio es aplicable a todas las categorías. Simplemente se les asigna el género o número que presenten en cada contexto. En este sentido nuestro etiquetario tiene una clara orientación a la anotación de corpus, frente a la marcación de información léxica. [Aguirre Moreno et al., 2001: 15]

O SLI-CTAG, pola súa banda, é congruente no tratamento das formas pronominais. En primeiro lugar, o valor *cortés* está reservado exclusivamente para *vostede/vostedes* e inclúense, como é lóxico, nos pronomes de *terceira* persoa. Logo, as caracterizacións son morfolóxicas, non sintácticas, é dicir, a etiqueta que se lle asigna a cada pronome correspóndese cos valores que ese pronome asume na teoría, como forma illada, non a que lle correspondería polo contexto. Vémolos máis claramente cuns exemplos. O pronome átono *nos* pode funcionar como acusativo, dativo ou ser un formante lexemático; pois ben, no SLI-CTAG aglutínanse nunha única etiqueta os tres valores, co que en realidade non se especifica o caso, isto é, non se realiza desambiguación. O mesmo sucede coas formas tónicas que se poden empregar como rectas ou oblicuas, tipo *el*, *nós*, etc., unha única etiqueta ambigua engloba os distintos valores que en ningún momento se van especificar. Nesta liña, o pronome máis complicado de todos por excelencia, *se*, etiquétase como *reflexivo*:

nos	P1CPC0	[Pronome, primeira, común, invariable, acusativo/dativo/reflexivo, non especificado]
nós	P1CPB0	[Pronome, primeira, común, invariable, nominativo/recto ⁷¹ , non especificado]
se	P3CNR0	[Pronome, terceira, común, invariable, reflexivo, non especificado]

En XIADA, como vimos anteriormente na descrición da categoría Pronome, realízase unha caracterización morfosintáctica detallada e, na anotación de corpus, sempre que o contexto nolo permite, especificase o valor casual, xenérico e numérico correspondente, coa excepción do pronome *se*, para o cal decidimos empregar unha etiqueta xeral debido ás dificultades que presentaba a súa desambiguación, tanto manual como automática. Un exemplo clarificador da análise pormenorizada que se intenta realizar en XIADA constitúeo a seguinte etiqueta, pertencente á análise do pronome *nos*:

O cambio é revolucionario pero pensa que aínda queda moito por camiñar polo que "non podemos sentirnos satisfeitas nin quedar na autocompracencia". [CORGAetq: ANT2008-07-10/75]

nos Raa1fp (Pronome, átono, acusativo, 1ª, feminino, plural)

A adaptación ao galego da clase Numeral de EAGLES realízase de forma diversa nos tres etiquetarios galegos.

Así, o SLI-TECTRA trata os numerais como un subtipo da clase Pronome, aínda que ao facer a correspondencia co etiquetario intermedio considéranos unha categoría independente. Polo que respecta á súa caracterización morfolóxica, distinguen entre *cardinais* e *ordinais* e aplican sistematicamente os valores de xénero (*masculino*, *feminino*) e de número (*singular*, *plural*) só no tipo *ordinal*, pois nos *cardinais* –deducimos pola información que se nos proporciona– unicamente as formas que presentan variación formal de xénero ou número son individualizadas: *un*, *unha* caracterízanse como masculino e feminino singulares e *dous*, *dúas* como masculino e feminino plurais; para os demais numerais (tipo *tres*) non se especifica ningún valor xenérico ou numérico:

⁷¹ Creemos que se produciu un erro e que no seu lugar querían indicar *nominativo/oblicuo*, xa que o uso dese valor se repite en todas as formas pronominais que poden funcionar como rectas e oblicuas (*nós*, *nosoutros*, *vostede(s)*, *vós*, *el(es)*, *ela(s)*). Ademais, o valor *recto*, illado, non está presente na táboa de valores, no seu sitio emprégase *nominativo* e *oblicuo* si o está.

tres PNC (*pronome numeral cardinal*)

O SLI-CTAG, pola súa banda, delimita os numerais nunha clase independente e distingue neles entre *cardinais*, *ordinais* e *partitivos*. Introduce nos atributos pertinentes, ademais de *xénero* e *número*, o *grao*, cuxos valores son *apreciativo* e *non especificado* e caracteriza os tres cardinais cos que exemplifica (*un*, *unha*, *tres*) co valor numérico *invariable*. Non sabemos se estamos ante un erro⁷² ou se, pola contra, todos os *cardinais* se definen no número con ese valor. Polo que respecta ao xénero, este só se especifica naqueles elementos que presentan formas diferenciadas para masculino e feminino; nos demais casos asígnaselles o valor *común*. Exs.:

un	MCOMN [<i>numeral, cardinal, non especificado, masculino, invariable</i>]
tres	MC0CN [<i>numeral, cardinal, non especificado, común, invariable</i>]
primeiriños	MOAMP [<i>numeral, ordinal, apreciativo, masculino, plural</i>]
medio	MPOMS [<i>numeral, partitivo, non especificado, masculino, singular</i>]

Como xa comentamos máis arriba para a clase Cifra, o sistema de XIADA, fronte ao do SLI-CTAG, trata do mesmo xeito na categoría Numeral a expresión da cantidade representada nuns casos mediante letras e noutros mediante cifras. Só se recoñecen os tipos *cardinal* e *ordinal*, pois os partitivos e multiplicativos intégranse nas clases Substantivo ou Adxectivo, segundo corresponda e, así mesmo, é o único que dá conta da función determinante ou nuclear dos seus elementos e que, ademais, sempre que o contexto o permite, individualiza os valores de xénero e número, asignándoos segundo se manifesten nas concordancias cos demais elementos do contexto. Exs.:

A extensión media das parcelas privadas é de unha ou dúas hectáreas, o que provoca que a xestión forestal non exista. [CORGAetq: ANT2005-03-03/57]

unha Ncdfs (*Numeral, cardinal, determinante, feminino, singular*)

dúas Ncdfp (*Numeral, cardinal, determinante, feminino, plural*)

En Galicia o paro descendeu en 900 persoas nos tres primeiros meses do ano, e en 14.500 nos últimos doce meses, co que hai 188.700 parados na comunidade, o 17,1% dos activos. [CORGAetq: CG1999-05-20/28]

900 Ncdfp (*Numeral, cardinal, determinante, feminino, plural*)

⁷² Non aparece ningún outro numeral cardinal nos fragmentos ilustrativos de anotación nin se explica ao longo do artigo que criterios se definiron para aplicar os distintos valores atributivos.

tres	Ncdmp (<i>Numeral, cardinal, determinante, masculino, plural</i>)
primeiros	Nodmp (<i>Numeral, ordinal, determinante, masculino, plural</i>)
14.500	Ncnfp (<i>Numeral, cardinal, non determinante, feminino, plural</i>)
doce	Ncdmp (<i>Numeral, cardinal, determinante, masculino, plural</i>)
188.700	Ncdmp (<i>Numeral, cardinal, determinante, masculino, plural</i>)
17,1	Ncnms (<i>Numeral, cardinal, non determinante, masculino, singular</i>)

Para a categoría Verbo, Aguirre Moreno et al. (2002a: 25-27; 2002b: 210, 212-215) afirman que tiveron que engadir atributos ou valores porque estes non se recollían na proposta de EAGLES. En concreto, marcan como innovación na táboa correspondente á xerarquía de subcategorías verbais os modos *infinitivo conxugado* e *xerundio conxugado*, o valor temporal *antepretérito* e no atributo *persoa a segunda cortés*. É verdade que na proposta inicial de Leech e Wilson (1996) non aparecen tales atributos, pero si se dá conta, meses máis tarde en Monachini e Calzolari (1996), do valor temporal pluscuamperfecto, do recoñecemento das segundas persoas cortesas e mais da caracterización do infinitivo conxugado na aplicación ao portugués do modelo GENELEX.

En teoría, excepto eses valores necesarios para a análise do galego, seguen a proposta de EAGLES sen outros cambios; non obstante, no inventario de etiquetas verbais (92 en total) obsérvanse anomalías. Por exemplo, a distinción *forma verbal finita* fronte a *forma verbal non finita* non se aplica, a pesar de recoñecela como un atributo; o *condicional*, recoñecido como un valor modal na táboa definatoria de atributos verbais, tradúcese no inventario en “indicativo pospretérito”; a denominación tradicional “imperfecto” do tempo verbal das táboas de subcategorías verbais e atributos e valores verbais convértese no inventario en “copretérito” e, finalmente, volven observarse no etiquetario do SLI-TECTRA etiquetas verbais de distinta lonxitude, o que implica que non todas as etiquetas verbais responden aos mesmos atributos. Exs.:

coller: VINF (*infinitivo*)

VINFCS2C (*infinitivo conxugado segunda singular de cortesía*)

VINFCS3 (*infinitivo conxugado terceira singular*)

En principio, os atributos e valores definidos no SLI-CTAG para a categoría verbal non presentan grandes diferenzas cos de XIADA: os valores modais, temporais,

xenéricos e numéricos coinciden na súa totalidade. O atributo *persoa* só difire porque en XIADA, a maiores, dámoslle cabida ao valor ambiguo “a” (*primeira/terceira*) xa que é probable que nalgunha ocasión non poidamos desfacer a ambigüidade existente entre a primeira e a terceira persoa do singular nas formas verbais dos tempos de indicativo copretérito, pospretérito e antepretérito, e no presente, pretérito e futuro de subxuntivo, así como no infinitivo conxugado. Por exemplo, a forma *cante* etiquétase:

SLI-CTAG	XIADA
VSP1S00 [<i>Verbo, subxuntivo, presente, 1ª, singular, xénero non especificado, grao non especificado</i>]	Vps10s (<i>Verbo, presente, subxuntivo, 1ª, xénero non aplica, singular</i>)
VSP3S00 [<i>Verbo, subxuntivo, presente, 3ª, singular, xénero non especificado, grao non especificado</i>]	Vps30s (<i>Verbo, presente, subxuntivo, 3ª, xénero non aplica, singular</i>)
VM03S00 [<i>Verbo, imperativo, tempo non especificado, 3ª, singular, xénero non especificado, grao non especificado</i>]	Vpsa0s (<i>Verbo, presente, subxuntivo, 1ª/3ª, xénero non aplica, singular</i>)

Tanto en XIADA como no SLI-CTAG o atributo *xénero* só é pertinente no participio e, ademais, nas formas de infinitivo e xerundio non conxugados, non aplican os atributos de *tempo*, *persoa*, *xénero* e *número*.

As únicas dúas diferenzas nos atributos entre estes dous últimos etiquetarios son de distinto calibre. A primeira delas non reviste maior importancia, pois non afecta á caracterización das formas verbais. Trátase da incongruencia que supón no SLI-CTAG situar na cadea de etiquetaxe antes o número ca o xénero. Non se explica por que se tomou esa decisión, pois nas demais categorías a orde é a inversa, pero supoñemos que se debe a que, excepto no participio, non é pertinente e que preferiron darlle prioridade posicional aos valores pertinentes na maior parte do paradigma. Tampouco ten relevancia a diferente escolla na orde dos atributos *tempo* e *modo*: no SLI-CTAG optaron por *modo* e *tempo* e en XIADA preferimos *tempo* e *modo*.

A segunda diferenza si afecta á caracterización morfolóxica da forma verbal e esta débese ao recoñecemento no SLI-CTAG do atributo *grao*, presente para etiquetar as formas do participio ou xerundio que leven o sufixo apreciativo:

cantandiño VX0000A [*Verbo, xerundio, tempo non especificado, persoa non especificada, número non especificado, xénero non especificado, apreciativo*]

cantadiño VP00SMA [*Verbo, participio, tempo non especificado, persoa non especificada, singular, masculino, apreciativo*]

En XIADA as formas verbais que leven o sufixo *-iño* non se caracterizan morfosintacticamente de xeito distinto; é dicir, non se marca a presenza do apreciativo; o tratamento destas formas é semellante ao descrito con anterioridade para os adverbios ou indefinidos con sufixo apreciativo ou superlativo.

Máis importancia nos parece que acada o feito de incluír no imperativo formas do presente de subxuntivo. O SLI-CTAG incorpora na táboa de exemplos de verbos para o Imperativo as formas *cante, cantemos* e *canten* para a 3ª persoa do singular e 1ª e 3ª do plural, respectivamente. Sorpréndenos esta decisión cando a fonte na que se fundamentan para a elaboración do etiquetario –Civit 2003– non inclúe as formas do Imperativo nos seus exemplos nin indica cantas formas contén este; ademais, no paradigma regular das *Normas* (RAG/ILG, 1995: 107; 2003: 106) só aparecen as segundas persoas do singular e do plural e en Álvarez e Xove (2002: 296), por citar unha gramática do galego recente, afirman que o Imperativo “*só conta cunha unidade non irreal con dous MPN, P2 e P5, con distribución moi reducida*”; quizais, a inclusión no Imperativo desas persoas veña determinada pola consideración en Álvarez, Regueira e Monteagudo (1986: 362) do “*imperativo como un subxuntivo reforzado, e de feito, atópase en distribución complementaria coas formas inequivocamente subxuntivas, no estilo indirecto e nas oracións negativas*” ou pola descrición de Freixeiro (2000: 370) baseada na *Gramática Galega* (1986: 362). En realidade a cuestión é polémica, porque uns autores consideran que o paradigma do imperativo debe ser ampliado con formas do subxuntivo, mentres que outros entenden que esas expresións non deixan de ser subxuntivas nestes novos usos, de xeito que os enunciados que constrúen non son propiamente imperativos. Non é a nosa intención tratar de solucionar un problema no que os estudosos non se poñen de acordo; simplemente queremos deixar constancia da diverxencia que se produce entre o etiquetario do SLI-CTAG e o de XIADA na consideración e delimitación das formas e persoas pertencentes ao Imperativo.

Finalmente, e xa para rematar a análise contrastiva que estamos levando a cabo nos etiquetarios existentes para a etiquetación de corpus en galego, só nos queda dar conta das categorías Substantivo e Adxectivo.

O SLI-TECTRA e o SLI-CTAG coinciden nos atributos considerados para a clase Nome, denominación que ambos prefiren sobre a de substantivo. A diferenza xorde nos valores delimitados e na súa aplicación.

Así, no SLI-TECTRA non se recollen os valores *común* e *invariable*, nin nada que se lles pareza, para caracterizar as palabras que formalmente non presentan oposición de xénero (*cantante*) ou de número (*martes*); asígnaselles o xénero e número que presenten en cada contexto. O problema que nós lle vemos a esta decisión xorde naqueles casos en que o contexto non sexa claro, que valor se escolle entón? Nos nomes propios sempre aplica o xénero e o número aínda que haxa casos difíciles de etiquetar como *París* ou *Internet*. A pregunta que nos facemos é que criterios se seguen para esa asignación? Finalmente, no SLI-TECTRA engádese o atributo *forma* cuxos valores son *plena* e *diminutivo*. Ás obxeccións que acabamos de sinalar, podemos engadir unha vez máis, na clase Nome, a non aplicación na cadea de etiquetaxe de todos os atributos implicados na caracterización dos substantivos. Ex.:

can NCMS (*nome común masculino singular*)

canciño NCDMS (*nome común diminutivo masculino singular*)

Ancares NPMP (*nome propio masculino plural*)

Burgas NPPF (*nome propio feminino plural*)

O SLI-CTAG non especifica nos propios os valores de *xénero*, *número* e *grao* e mantén a atribución ambigua naquelas formas que presentan sincretismo formal no *xénero* ou no *número*. Exs.:

xornalista CMS [*Nome, común, común, invariable*]

neniño NCMSA [*Nome, común, masculino, singular, apreciativo*]

Breogán NP000 [*Nome, propio, xénero non especificado, número non especificado, grao non especificado*]

En XIADA, pola contra, a semellanza do realizado noutras categorías, non recoñecemos o atributo *grao*, e desambiguamos os atributos de *xénero* e *número* sempre que o contexto nolo permite, especificando tamén o valor *masculino*, *feminino*, *singular* e *plural* nas formas que presentan sincretismo formal para estes atributos. Nos propios, os criterios que rexen a asignación dos valores xenéricos e numéricos son máis complexos:

a) Nos topónimos, por exemplo, consideramos que non é pertinente nin o xénero nin o número porque dependendo do xenérico no que pensemos (poboación, cidade, pobo, lugar, etc.) a concordancia variará.

b) Nos antropónimos distinguimos entre femininos (*María, Teresa Vázquez*) e masculinos (*Pedro, Manuel Fernández*), pero neles, por regra xeral, non aplicamos o número a non ser “*cando pasan a usarse á maneira de adxectivos, ben para referirse a toda unha familia* (os Andrades), *ben usados metonimicamente* (Seica ten tres Laxeiros) *ou ben como calificativos atribuíndolle a un individuo as cualidades que se recoñecen á persoa portadora do nome propio* (Estamos un pouco faltos de Castelaos e de Bóvedas)” (Álvarez, Regueira e Monteagudo, 1986: 66).

c) Nos nomes de *organismos, empresas, marcas, etc.* o contexto e as concordancias nel expresadas determinarán a aplicación ou non do xénero e o número.

Exemplos de etiquetación de Substantivos:

Cal é o perfil do xornalista galego? [CORGAetq: ANT2005-03-03/28]

xornalista Scms (*Substantivo, común, masculino, singular, (lema) xornalista*)

Os resultados veñen acompañados dunha simulación dunha loita entre dous homiños. [CORGAetq: DLV2007-05-15/53]

homiños Scmp (*Substantivo, común, masculino, plural, (lema) home*)

O axudante do adestrador do combinado de baloncesto foi Rubén Domínguez, segundo no propio Breogán. [CORGAetq: ANT2008-07-10/87]

Rubén Domínguez Spm0 (*Substantivo, propio, masculino, número non aplica, (lema) Rubén Domínguez*)

Breogán Spms (*Substantivo, propio, masculino, singular, (lema) Breogán*)

Na clase Adxectivo, o SLI-TECTRA mantén os mesmos atributos e valores ca na clase Nome, cos engadidos do *grao* (*positivo, comparativo e superlativo*) e do valor *apocopado* no atributo *forma*. Exs.:

novo AXMS (*adxectivo masculino singular*)

noviño AXDMS (*adxectivo diminutivo masculino singular*)

novísimo AXSMS (*adxectivo superlativo masculino singular*)

gran AXAPMS (*adxectivo apocopado masculino singular*)

AXAPFS (*adxectivo apocopado feminino singular*)

O SLI-CTAG, excepto no tipo *común vs. propio*, o cal non é pertinente no Adxectivo, mantén os mesmos atributos e valores que para a categoría Nome: *grao* (*apreciativo*), *xénero* (*masculino, feminino, común*) e *número* (*singular, plural, invariable*). Sorprendentemente, o valor *apreciativo* engloba o grao comparativo e superlativo co que, neste etiquetario e no corpus con el etiquetado, non é posible consultar ou extraer unicamente os comparativos e superlativos sintéticos. Exs.:

ecolóxico	A0MS [<i>Adxectivo, grao non especificado, masculino, singular, (lema) ecolóxico</i>]
feble	A0CS [<i>Adxectivo, grao non especificado, común, singular, (lema) feble</i>]
choromicas	A0CN [<i>Adxectivo, grao non especificado, común, invariable, (lema) choromicas</i>]
grandiño	AAMS [<i>Adxectivo, apreciativo, masculino, singular, (lema) grande</i>]

Por último, en XIADA, os adxectivos son caracterizados cos atributos *grao* (comparativo, superlativo, non aplica), *xénero* (masculino, feminino, masculino/feminino) e *número* (singular, plural, singular/plural). Exs.:

Fóra da imaxe de feble saúde, a secuestrada falou varias horas e iniciou decontado unha xira que a levou a Francia e que pronto a deixará ver en Madrid ou Roma, por exemplo. [CORGAetq: ANT2008-07-10/82]

feble A0fs (*Adxectivo, grao non aplica, feminino, singular, (lema) feble*)

Se somos capaces de darmos unha educación para a información. [CORGAetq: ANT2005-03-03/75]

capaces A0ap (*Adxectivo, grao non aplica, masculino/feminino, plural, (lema) capaz*)

Por eso un porco ibérico de belota é de manexo caro e de alimentación carísima. [CORGAetq: CG2001-10-10/35]

caro A0ms (*Adxectivo, grao non aplica, masculino, singular, (lema) caro*)

carísima Asfs (*Adxectivo, superlativo, feminino, singular, (lema) caro*)

O homenaxeado era un tal Rogelio O Pichón, un imbécil, un coitadiño con ínfulas de toureiro, que se deixou convencer para un esperpéntico retrato con estoque, capote e traxe de luces. [CORGAetq: ANT2008-07-10/46]

coitadiño A0ms (*Adxectivo, grao non aplica, masculino, singular (lema) coitado*)

A partir da análise comparativa que acabamos de realizar podemos colixir que, dos tres etiquetarios existentes para a anotación de corpus en galego, o que ofrece máis

prestacións é o *tag-set* de XIADA, non só pola análise pormenorizada que proporciona para cada clase de palabra, senón tamén pola evidente orientación á etiquetación e desambiguación de corpus. A detallada caracterización morfosintáctica das unidades léxicas non é tara para a consulta dun corpus etiquetado co etiquetario de XIADA; todo o contrario, permite tanto consultas xenéricas, nas que podemos prescindir dos atributos que nun momento dado non nos interesen, como consultas moi específicas. Debido á súa estruturación e detalle de análise, é posible ir baixando chanzos (*categoría, tipo, valor, xénero, etc.*) ata chegar ao nivel de análise máximo, a etiqueta completa; noutras palabras, se a información figura no corpus podemos escoller o nivel de análise que nos interese; se, pola contra, a información non está caracterizada xa non temos opción, a escolla é imposible.

1.6. Conclusións

Son varias as conclusións que podemos extraer despois de describir as recomendacións de EAGLES para construír un etiquetario, ver como estas se aplicaron en linguas próximas á nosa, definir o etiquetario de XIADA e contrastalo cos outros dous existentes para o galego.

En primeiro lugar, hoxe en día, a falta de homoxeneidade segue sendo unha constante no tratamento dos elementos que poden funcionar como determinantes ou como núcleos. Así, podemos establecer a seguinte tipoloxía:

A) Etiquetarios nos que se mantén unha dobre etiquetación. Un exemplo claro desta modalidade constitúeo o *tag-set* do corpus CLiC-TALP (Civit et al., 2001) que inclúe os demostrativos, posesivos, interrogativos, indefinidos, cardinais e ordinais como tipos das clases Determinante e Pronome. Outro caso máis é o do Corpus AnCora⁷³ coa particularidade de que neste non se diferencia entre cardinais e ordinais, senón que se agrupan no tipo numeral.

B) Etiquetarios nos que se recorre a unha categoría especial para acoller os elementos que poden funcionar como adxectivais ou pronominais. É o que sucede coa clase Especificador do *tag-set* do IULA (Morel et al., 1997 e 1998), creada para indicar

⁷³ Dispoñible en http://clic.ub.edu/ancora/lng/en/morfologia_tagset.pdf [consultado o 18/02/2010]

que nunha etapa posterior –suponse que desambiguación– se decidirá se a unidade en cuestión é un pronome ou un adxectivo.

C) Etiquetarios nos que non se prevé empregar a distinción funcional entre pronome e determinante. O *tag-set* do *Corpus Anotado do Português Europeu*⁷⁴ recoñece as categorías Demostrativo, Indefinido, Posesivo, Numeral e agrupa nunha única clase Relativo, Interrogativo e Exclamativo. Os dous etiquetarios do Seminario de Lingüística Informática da Universidade de Vigo (SLI), o aplicado ao corpus TECTRA (Aguirre Moreno et al., 2001, 2002a, 2002b) e o usado para o CTAG (Gómez Guinovart e López Fernández, 2009), constitúen outros dous representantes deste subtipo de etiquetarios, coa diferenza de que o primeiro considera subtipos do Pronome o que o segundo delimita como clases categoriais distintas: demostrativos, posesivos, indefinidos, interrogativos, relativos e numerais.

D) Etiquetarios, finalmente, coma o de XIADA que distinguen entre os usos determinantes ou pronominais dos elementos das categorías Demostrativo, Posesivo, Indefinido, Interrogativo-Exclamativo, Relativo e Numeral.

En segundo lugar, podemos constatar a validez e adecuación do etiquetario empregado en XIADA na análise morfosintáctica do galego actual, pois este foi aplicado xa a un subconxunto de textos do CORGA, o CORGAetq, cuxa etiquetación se pode consultar de xeito gratuito e libre no enderezo <http://corpus.cirp.es/corgaetq>.

Así mesmo, probamos que empregando un etiquetario de máis de 100 etiquetas, cifra máxima recomendada na proposta de EAGLES para a anotación de corpus (Monachini e Calzolari, 1996), en concreto 299 ata o momento, obtemos porcentaxes de acerto semellantes ás conseguidas para linguas coma o inglés (95%-97%), pois estamos no 95,99% de eficacia.

En terceiro lugar, despois de inventariar as etiquetas que se empregaron na desambiguación do CORGAetq e as que figuran no núcleo de XIADA (Apéndice B e C), comprobamos a necesidade de ampliar o corpus de adestramento con outro tipo de documentos, narrativa por exemplo, que conteña máis diálogo, entrevista, etc. e menos

⁷⁴ Disponible en http://www.clul.ul.pt/sectores/linguistica-de-corpus/manual_annotacao.pdf [consultado o 12/03/2010]

crónica ou sucesos, co que é probable que bastantes das 84 etiquetas que faltan por documentar en texto real aparezan.

En cuarto lugar, as 70 etiquetas recollidas unicamente no lexicón non se deben ter en conta no cómputo xeral de etiquetas nin deben ser reflectidas no etiquetario, posto que non ten razón de ser manter un valor atributivo xenérico que nunca pode aparecer en texto anotado e cuxa orixe foi, e segue a ser, a de facilitar a introdución na base de datos da caracterización morfosintáctica de certos elementos das clases pechadas ou dos adverbios.

Finalmente, e en quinto lugar, destacamos a marcada orientación do *tag-set* de XIADA cara á etiquetación e desambiguación de corpus textuais, non presente noutros etiquetarios do galego, pois non só se especifica, sempre que o contexto o permite, o xénero e número de substantivos e adxectivos que presentan sincretismo formal nestes atributos, senón que tamén se indica, por exemplo, a función que desempeñan os tradicionalmente denominados determinativos ou os adverbios.

Capítulo 2

Elaboración da gramática formal do galego actual e a súa implementación nun lexicón.

2.1. *Introdución*

É obvio que para deseñar un etiquetario pormenorizado coma o de XIADA non podemos prescindir das palabras que integran cada unha das categorías gramaticais, senón que debemos ter en conta todas as peculiaridades lingüísticas que as caracterizan. Témolas en mente para decidir que atributos aplican en cada caso e cales son os seus valores, mais dun xeito abstracto; non traballamos con elas e si coa súa análise morfosintáctica. Decidido cal vai ser o etiquetario que empreguemos na anotación de corpus, chega o momento de dar outro paso; é hora de acometer a elaboración do lexicón, fase na que nos centraremos neste capítulo e na que as palabras asumen o protagonismo.

Segundo o *Gran Dicionario Xerais da Lingua* (Carballeira Anllo, 2000, 2009)⁷⁵, un lexicón é un “léxico ou dicionario, especialmente dunha lingua clásica”, e de acordo co *Dicionario de galego* de Ir Indo⁷⁶ trátase dun “compoñente da gramática que, segundo a lingüística xenerativa transformacional, describe a competencia léxica. Consta xeralmente dun conxunto de entradas léxicas, que se especifican fonolóxica, semántica e sintacticamente, e dun número variable de regras”.

No campo da lingüística computacional, Schiller e Karttunen aclaran:

The purpose of a tagger lexicon is to provide for each *token* of a text its potential properties. What this means, exactly, depends on the tagging system of which the lexicon is part. At the input side of the lexicon component, the tokens which correspond to lexicon entries must be compatible with the tokenizer output. This also means, e.g., that if the tokenizer recognizes multi-word units, the lexicon should cover them as well. [Schiller e Karttunen, 1999: 135]

⁷⁵ En diante, Xerais.

⁷⁶ Dispoñible en <http://www.digalego.com> [consultado o 07/06/2010]. En diante, Ir Indo, tanto para a versión electrónica como para a de papel.

É dicir, un lexicón destinado a etiquetar corpus debe prover a cada *token* do texto das súas características potenciais e debe, ademais, concordar no seu contido coas propiedades do segmentador de unidades ou *tokenizador*.

Seguindo a Schiller e Karttunen (1999: 135-147), o tipo máis simple de lexicón para a etiquetaxe de corpus indica para cada palabra, mediante etiquetas, a(s) categoría(s) á(s) que pertence o *token*, estando sempre determinado o lexicón pola elección do etiquetario. Non obstante, un lexicón, continúan, pode incluír información adicional sobre a flexión do lema ou lemas do que a palabra é mostra. Examinando o contido do lexicón e a estrutura empregada para representalo, distinguen tres tipos:

1) *Full-form lexicons*. Neles, as formas flexionadas emparéllanse cunha ou máis etiquetas que indican a clase de palabra á que pertence. As entradas teñen a seguinte estrutura:

<palabra> [etiqueta clase de palabra]

2) *Morphological lexicons*. O lexicón posúe información sobre o lema e os trazos morfolóxicos de cada palabra. A súa estrutura inclúe

<palabra> [(<lema> ([etiqueta morfolóxica]+) <etiqueta clase de palabra>]

3) *Weighted lexicons*. A información estatística (<P>) está asociada no lexicón a cada etiqueta:

<palabra> [<etiqueta clase de palabra> (<P>)]

Como acabamos de ver, un etiquetador necesita consultar o lexicón onde se almacena, como mínimo, a información de etiqueta que lle corresponde a cada unidade léxica pois, aínda que en xeral os etiquetadores intentan adiviñar a etiqueta dunha forma gráfica que non estea presente no lexicón, canto maior sexa este máis posibilidades de acerto hai.

Á hora de afrontar a elaboración dun recurso léxico e gramatical calquera destinado a un uso computacional é fundamental que no seu deseño se cubran todas as necesidades actuais e, ao tempo, se prevexan as futuras. É primordial, polo tanto, decidir que tipo de información lle queremos proporcionar a cada entrada e como imos implementar esa información. Ademais, debemos reflexionar sobre a súa utilidade, pois un lexicón non ten por que estar destinado unicamente a etiquetar e lematizar un texto.

Se é o suficientemente detallado e complexo pode servir de punto de partida para outras utilidades como a construción dun corrector ortográfico ou un conxugador de verbos.

No noso caso, o obxectivo que perseguimos lingüisticamente é que o lexicón de XIADA, ademais de conter a información que permita a identificación e caracterización morfolóxica plena de calquera palabra galega, nos proporcione información sobre a normatividade actual da entrada, a procedencia do seu lema e o ámbito léxico ao que pertence. A maiores, buscamos un sistema de implementación sinxelo, cómodo e intelixente que nos permita a introdución das distintas entradas en bloques e sen necesidade de ter que incorporar a unidade completa en cada caso.

2.2. O lexicón de XIADA

2.2.1. Estrutura xeral

Antes de nada gustaríanos aclarar que o lexicón de XIADA, a pesar de clasificarse entre os lexicóns morfolóxicos de Schiller e Karttunen (1999: 135-147), non está constituído por unha mera relación ou listaxe de formas gráficas ás que se lles proporciona a etiqueta e lema correspondentes. Aquí, en xeral, non hai formas; estas teñen que ser xeradas a partir da información que contén cada entrada. Por exemplo, a unidade *nenas* non aparece como tal en ningún lugar do lexicón. É recoñecida polo etiquetador e xerada polo sistema porque a entrada *nen*, cuxo lema e subetiqueta son respectivamente *nen* e Sc (*Substantivo común*), remítese ao grupo de derivación G1, onde figura unha desinencia “as” coa caracterización morfolóxica de *feminino plural*.

Isto débese a que, dado que perseguíamos que a introdución e xestión dos datos fose práctica e nos permitise aforrar tempo e espazo en disco, estimamos que maioritariamente as entradas do lexicón tiñan que ser caracterizadas morfoloxicamente desagregándoas en *lema*, *raíz*, *subetiqueta* e *grupo de derivación*. Con esta estrutura evitamos a introdución, por exemplo, de 70 formas verbais por cada un dos verbos que integramos no lexicón; aínda que, en contrapartida, esixiunos un estudo formal detallado dos integrantes de cada unha das clases gramaticais variables para agrupalos en modelos, e deste xeito construír a gramática formal que nos permitise analizar e tamén xerar as formas flexionadas e conxugadas do galego moderno.

Todas as entradas do lexicón constan, pois, de lema, raíz, subetiqueta e, maioritariamente nas pertencentes ás clases flexivas, de grupo de derivación. Malia que son termos bastante estendidos en lingüística computacional é conveniente aclarar que entendemos por cada un deles, posto que os imos empregar con frecuencia neste capítulo. Así:

1) O lema é o representante das distintas formas flexionadas ou conxugadas que se integran nun paradigma e, obrigatoriamente, cada entrada do lexicón debe remitirse a un lema. No noso sistema, ademais, a clase morfolóxica é un elemento constitutivo do lema, polo que todo lema está asociado a unha clase de palabra. Así, para *galego*, por exemplo, teremos dous lemas: un, asociado á categoría substantivo e outro á de adxectivo.

2) Basicamente, podemos considerar que a raíz é o lema sen morfemas gramaticais. Por exemplo, ao lema *galego* correspóndelle a raíz *galeg*. Raíz e lema coinciden nas entradas das clases invariables e nas dos substantivos e adxectivos que só presentan flexión no número. En ningún caso se pode identificar esta raíz co lexema da gramática tradicional, posto que para o seu establecemento non temos en conta os morfemas derivativos, senón só os flexivos.

3) O termo subetiqueta empregámolo para referirnos á caracterización morfolóxica básica da entrada no lexicón. Continuando a exemplificación con *galego*, as subetiquetas correspondentes ás entradas como substantivo e adxectivo son, respectivamente, Sc (*Substantivo, común*) e A0 (*Adxectivo, grao non aplica*).

4) Finalmente, os grupos de derivación constitúen conxuntos de terminacións para as que se proporcionan uns valores. Por exemplo, as dúas raíces *galeg*, pertencentes aos lemas *galego* substantivo e *galego* adxectivo, remítense ao grupo de derivación G1 no que se interpreta que se a entrada

remata en	<i>o,</i>	a caracterice como =>	<i>masculino singular</i>
"	<i>a,</i>	" =>	<i>feminino singular</i>
"	<i>os,</i>	" =>	<i>masculino plural</i>
"	<i>as,</i>	" =>	<i>feminino plural</i>

Deste xeito, para todos os substantivos e adxectivos que teñen flexión de xénero e forman o plural engadindo *-s*, tipo *nenho*, introducimos só o lema e a raíz correspondente, sen necesidade de ter que inserir todas as formas do paradigma.

Indicabamos nos obxectivos desexados para o noso lexicón que, ademais de buscar a identificación e caracterización morfolóxica plena de cada unidade léxica, nos gustaría contar con información sobre a procedencia do lema, o ámbito léxico ao que pertence e a indicación sobre a normativa oficial actual. Obviamente non precisamos estes datos para etiquetar corpus, mais se dispoñemos deles sempre poderemos atoparlles utilidade; pola contra, se non os incorporamos cando construímos o lexicón, facelo máis tarde sería altamente custoso e, se non os incluimos, en ningún momento poderemos dispoñer desa información. Por riba, estaríamos perdendo unha boa oportunidade de contar cun lexicón amplo e completo no que retratar a realidade lingüística actual do galego.

Pois ben, no lexicón de XIADA, á parte de lema, raíz, subetiqueta, categoría gramatical e grupo de derivación, cada entrada responde aos parámetros seguintes: indicación sobre a normativa oficial, ámbito léxico e procedencia do lema.

a) Normativa oficial. Para un correcto recoñecemento das formas existentes nos textos galegos contemporáneos necesitamos que o lexicón sexa amplo e que non só estea integrado por lemas normativos. Se este fose o caso, moitos dos documentos do CORGA non poderían ser etiquetados ou serían a medias. É fundamental, polo tanto, darlles cabida tamén a dialectalismos, hiperenxebrismos, castelanismos, etc., porque estes aparecen nos textos. Debemos, así a todo, poder diferenciarlos, e por iso introducimos este criterio na caracterización de lemas, formas e incluso algunha desinencia. Por que tamén nas formas? Basicamente, porque a normatividade do lema non implica a normatividade da forma. Por exemplo, *fosemos*, forma totalmente correcta ata xullo de 2003 en que entrou en vigor a actual normativa, remítese ao lema *ser* ou *ir*, normativos; non obstante, *per se*, é un *token* que responde negativamente á normativa actual.

A economía lingüística e a reiteración impoñennos a obriga de aplicar este parámetro a algunhas desinencias, xa que caracterizan formas dialectais recorrentes nos textos. É bastante frecuente, por poñer un exemplo, para o plural das palabras rematadas

en *-n*, a aparición de formas co plural en *-s* no sitio do normativo *-ns*: *opiniós, razóns*, etc. en vez de *opinións, razóns...* Coa estrutura que empregamos en XIADA a opción máis rendible é a de caracterizar a desinencia *-s* no seu valor de *p* (*plural*) como non normativa no grupo de derivación G46 ao que se remiten 2957 entradas, as cales rematan en *-n* e forman o plural incorporando *-s* (normativo) ou prescindindo da consoante nasal e engadindo *-s* (non normativo).

b) *Ámbito léxico*. Cremos que pode ser útil diferenciar na estrutura do lexicón módulos segundo o tipo de léxico que conteñan que poidan ser combinados. Para iso dispuxemos no deseño da estrutura, e obviamente implementamos na construción do lexicón, dun campo máis no que se caracteriza o lema como pertencente ao léxico común ou ao técnico-científico, especificando neste último o ámbito no que se clasifica: administración, economía, medicina, etc.

Polo momento, o noso sistema conta soamente co lexicón *xeral*, no que se inclúe o léxico común e cuxa descrición constitúe o núcleo deste capítulo, e mais o lexicón *médico*, no que se introduciu o vocabulario específico do ámbito da medicina. Este último non será obxecto do noso estudo pois, malia que coordinamos a súa elaboración, non nos ocupamos de implementalo⁷⁷.

c) *Fonte*. Nos textos actuais preséntanse formas que non aparecen nos dicionarios-vocabularios preceptivos da lingua galega, ben porque son termos técnicos para os que se acaba de propoñer unha denominación, ben porque se documentan cun uso categorial distinto; non obstante, a súa introdución no lexicón é imprescindible para o recoñecemento e caracterización dos textos. Coa indicación da fonte achegamos datos concretos sobre a localización do lema e a fiabilidade da forma.

No noso lexicón a información referida á fonte aparece recollida na última columna da entrada e nela indícase a fonte do lema na que nos baseamos para a incorporación de cada unha das entradas. A correspondencia entre os códigos que poden aparecer nesa columna e a referencia bibliográfica que representa, organizada cronoloxicamente en orde decrecente, é a seguinte:

⁷⁷ O lexicón *médico* é resultado dun convenio entre a Secretaría Xeral de Política Lingüística e a Consellería de Sanidade, cuxo obxectivo era proporcionar o material lingüístico necesario para construír o corrector ortográfico galego do proxecto IANUS.

- corga:** Centro Ramón Piñeiro para a Investigación en Humanidades (2010): *Corpus de Referencia do Galego Actual, versión 1.5*, <<http://corpus.cirp.es/corga>>.
- xerais_2009:** Carballeira Anllo, X. M. (coord.) (2009): *Gran Dicionario Xerais da Lingua*, Vigo, Xerais.
- volga_2004:** González González, M. e A. Santamarina Fernández (coords.) (2004): *Vocabulario ortográfico da lingua galega (VOLGa)*, A Coruña / Santiago de Compostela, Real Academia Galega / Instituto da Lingua Galega.
- drag_2004:** García, C. e M. González González (drs.) (2004): *Dicionario castelán-galego da Real Academia Galega*, A Coruña, Fundación Pedro Barrié de la Maza.
- olg_2004:** González Rei, B.: *Ortografía da lingua galega*, A Coruña, Galinova Editorial.
- irindo_2004:** Ledo Cabido, Bieito (dr.) (2004): *Dicionario de galego*, Vigo, Ir Indo.
- normas_2003:** Real Academia Galega e Instituto da Lingua Galega (2003): *Normas ortográficas e morfolóxicas do idioma galego*, A Coruña / Santiago de Compostela, Real Academia Galega / Instituto da Lingua Galega.
- xerais_2000:** Carballeira Anllo, X. M. (coord.) (2000): *Gran Dicionario Xerais da Lingua*, Vigo, Xerais.
- drag_1997:** García, C. e M. González González (drs.) (1997): *Diccionario da Real Academia Galega*, A Coruña, Galaxia.
- normas_1995:** Real Academia Galega e Instituto da Lingua Galega (1995): *Normas ortográficas e morfolóxicas do idioma galego*, A Coruña / Santiago de Compostela, Real Academia Galega / Instituto da Lingua Galega.

gram_1986⁷⁸: Álvarez R., X. L. Regueira e H. Monteagudo (1986): *Gramática galega*. Vigo, Galaxia.

Finalmente, polo que respecta aos parámetros de tipo de léxico e fonte, consideramos que o *Vocabulario ortográfico da lingua galega* (VOLGa) (González e Santamarina, 2004), como texto sancionador máis recente e posuidor dun número de entradas aceptable, definirá os lemas pertencentes ao léxico común ou principal do lexicón de XIADA.

Para que a estrutura do lexicón sexa permeable e a identificación e asignación dos distintos modelos formais que o integran máis accesible, inserimos no deseño da estrutura un *representante de grupo* que serve de modelo para cada un dos grupos existentes en XIADA. Así o *representante de grupo* do grupo derivacional G1 é *nenó*.

Recapitulando, o lexicón de XIADA, ademais de conter os datos que posibilitan a identificación e caracterización morfolóxica plena de calquera unidade léxica galega actual, proporciona información sobre a normativa oficial, o ámbito léxico e a procedencia da forma.

Así, por exemplo, cando introducimos unha unidade coma *galego* no lexicón a información que proporcionamos é a seguinte:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	galeg	Sc	G1	galego	S	si	volga_2004
xeral	principal	galeg	A0	G1	galego	A	si	volga_2004

2.2.2. Implementación do vocabulario no lexicón

Sinalabamos anteriormente que a elección dun sistema de introdución do vocabulario galego no lexicón, cómodo e práctico, provocou a elaboración de modelos formais que desen conta das peculiaridades flexivas dos distintos vocábulos para agrupalos e sistematizar a súa introdución e caracterización no lexicón de XIADA. Isto implica crear paradigmas formais que acollan as raíces e grupos de derivación; é dicir, clasificar formalmente todas as posibles variacións non só á hora de flexionar o xénero ou o número, senón tamén tendo en conta as modificacións gráficas que sofre a raíz, sexan estas acentuais (*compás* vs. *compases*) sexan alternancias nos caracteres (*rapaz* vs. *rapaces*).

⁷⁸ Curiosamente é a única fonte na que documentamos o lema *lentes* como substantivo masculino plural, cuxo uso chega nos textos do CORGA ata o ano 2006, e o cal nas demais fontes lexicográficas consideradas aparece xa como substantivo feminino.

2.2.2.1. Categorías invariables

A introdución no lexicón dos elementos pertencentes ás clases de palabras invariables resulta bastante sinxela, posto que non é preciso, polas súas características, remitilas a ningún grupo de derivación; a raíz e mais o lema coinciden nas conxuncións, interxeccións, preposicións, locucións conxuntivas, locucións preposicionais e sinais de puntuación. Para todos os elementos integrantes destas clases a subetiqueta proporciona xa a caracterización morfolóxica completa. Un exemplo da información e estrutura que se seguiu para introducir os elementos conxuntivos, interxectivos, preposicionais e de puntuación podémolo ver na seguinte táboa:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	daquela	Cs		daquela	C	si	volga 2004
xeral	principal	con	P		con	P	si	volga 2004
xeral	principal	abur	Y		abur	Y	si	volga 2004
xeral	principal	;	Q;		;	Q	si	corga

Para as locucións conxuntivas e preposicionais, a maiores da información anterior, engadimos un campo máis no que reflectimos se a locución é ambigua ou non; é dicir, facemos constar no lexicón se a unidade multipalabra conxuntiva ou preposicional se analiza sempre como locución ou se pode presentar outra análise distinta. Así no campo *segura* aparecerá *si* ou *non* dependendo de cada caso concreto. Por exemplo⁷⁹:

locución	subetiqueta	grupo	lema	categoría	normativa	segura	fonte
sempre e cando	Lcs		sempre e cando	L	si	si	drag 1997
se ben	Lcs		se ben	L	non	non ⁸⁰	corga
cara a	Lp0		cara a	L	si	non	volga 2004
con relación a	Lp0		con relación a	L	si	si	drag 2004

Somos conscientes de que no inventario anterior de categorías invariables non incluimos o adverbio, malia selo. Esta circunstancia débese a dous feitos fundamentais: por un lado, certos adverbios admiten o sufixo apreciativo/intensificador *-iño* e o superlativo *-ísimo*; por outro, a existencia dos denominados adverbios en *-mente*.

⁷⁹ Por motivos de representación prescindimos na táboa exemplificativa dos campos *lexicón* e *sublexicón* que remiten a información, respectivamente, ao lexicón *xeral* e ao sublexicón *principal*.

⁸⁰ Malia que a primeira vista poida parecer unha locución segura, hai casos nos que debe analizarse cada elemento por separado coma, por exemplo, na seguinte oración na que corresponde a caracterización de conxunción e adverbio: *Se ben che ole, mellor che saberá!* (CORGA: T. Moure, *Unha primavera para Aldara*, Deputación Provincial da Coruña, 2008). O mesmo ocorre con *cara a*, para a que pode corresponder a análise de substantivo e mais preposición: *Decidira verlle a cara a Terencio.* (CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984).

O primeiro dos aspectos implica que raíz e lema non coinciden en todos os casos; e no segundo, ata o 2006, non se introducen no lexicón esas expresións, senón que se creou unha regra lingüística para detectalas e caracterizalas, sen que estivesen presentes na relación de raíces e lemas.

Como adiantabamos no capítulo primeiro, no etiquetario de XIADA non hai ningún atributo cuxo valor sexa apreciativo ou diminutivo, a pesar da frecuencia de uso deste no galego, pois, igual ca os despectivos e aumentativos, trátase de sufixos alterativos, ou sexa, a engádega deses sufixos non altera a categoría gramatical da base léxica, non se produce transcategorización. Todos eles achegan nocións semánticas, connotacións que decidimos non incluír no etiquetario e, polo tanto, tampouco na descrición morfolóxica, mais que si poden localizarse, tanto no lexicón como no CORGAetq, mediante a consulta combinada dos criterios *forma*, *etiqueta* e *lema* (fundamentalmente nas clases Substantivo e Adxectivo, pero tamén Indefinido, Verbo ou Adverbio).

A non individualización no etiquetario non significa que non introduzamos no lexicón formas diminutivas ou aumentativas. En XIADA tamén teñen cabida, só que o lema que lles asignamos é a forma base, sen o sufixo. Deste xeito, nas listaxes de lemas ou no cómputo destes as unidades diminutivas, aumentativas ou superlativas non figuran. Exs.:

lexicón	sublexicón	raíz	subetiqueta	lema	categoría	normativa	fonte
xeral	principal	cerquiña	Wn	cerca	W	si	volga 2004
xeral	principal	prontiño	Wn	pronto	W	si	volga 2004
xeral	principal	lonxísimo	Wn	lonxe	W	si	volga 2004
xeral	principal	a modiño ⁸¹	La0	a modo	L	non	volga 2004

Os adverbios en *-mente*, a efectos da súa incorporación ao lexicón, deberían figurar na mesma lista ca as preposicións ou conxuncións, non obstante, como xa adiantabamos máis arriba, non aparecen porque, en realidade, non se introduciu no lexicón ningún adverbio en *-mente* ata o 2005-2006.

Todos sabemos que un dos procedementos polos que se obteñen adverbios, principalmente modais, consiste na derivación por medio do sufixo *-mente* a partir dun

⁸¹ A consulta en XIADA das locucións, siglas, abreviaturas e símbolos realízase á parte de todos os demais tipos ou subtipos, polo que en vez de *raíz* o nome do campo é *locución* e, a maiores, tamén falta nesta táboa o campo *segura* que leva o valor *si* para a locución adverbial *a modiño*; non obstante, a efectos de exemplificación do sufixo *-iño* cremos que estes detalles non teñen maior importancia.

adxectivo na súa forma feminina singular. Pois ben, a difusión e o automatismo deste recurso levounos a pensar que era factible crear unha regra lingüística no sistema de XIADA de xeito que todas as palabras rematadas en *-mente*, excepto *clemente*, *demente*, *deprimente*, *dirimente*, *dormente*, *eximente*, *inclemente*, *mente*, *oprimente*, *remanente*, *semente*, *tremente* e *vehemente*⁸² se, prescindindo do sufixo *-mente*, coincidía o resto da palabra cunha forma cuxa etiqueta puidese ser adxectivo feminino singular, considerase o analizador morfolóxico que se trataba dun adverbio para o cal forma e lema eran idénticos e cuxa etiqueta tiña que ser *Wa (Adverbio, nuclear/modificador)*. Logo, xa o etiquetador se encargaría de decidir en cada caso concreto se o valor atributivo que correspondía era o *nuclear* ou o *modificador*.

Por razóns que non veñen ao caso, esta regra non se implementou no seu momento no código interno do analizador morfolóxico e, posteriormente, na negociación co responsable informático decidimos ignorala, fundamentalmente por cuestións de tempos, co que nos vimos abocados a introducilos á man, aínda que dun xeito semiautomático. Para iso extraemos todos os adverbios en *-mente* do VOLGa, 1716 lemas, e implementámoslos no lexicón.

Os adverbios, conxuncións e mais os tres tipos de locucións clasificados en XIADA, conxuntivas, preposicionais e adverbiais, foron introducidos manualmente no sistema a partir das listas presentes na edición das *Normas* de 1995 (RAG/ILG, 1995: 173-186). No 2005-2006, cando acometemos a adaptación do lexicón á normativa actual, extraemos sistematicamente as locucións preposicionais e adverbiais que aparecen no VOLGa⁸³ e introducimos no lexicón aquelas que faltaban. Ademais, conmutámoslles a todos os lemas a fonte primaria pola de *volga_2004*, sempre que estivesen nesa obra lexicográfica e coincidise a categoría gramatical asignada⁸⁴. A esas alturas xa había uns anos que, grazas á incorporación dun enxeñeiro informático ao

⁸² Relación facilmente conseguida a partir do *Diccionario inverso da lingua galega (Versión provisional)*.

⁸³ Debemos agradecer ao profesor Manuel González González a súa colaboración e axuda por proporcionarnos unha copia do VOLGa en formato electrónico que nos facilitou inmensamente o labor a realizar.

⁸⁴ Lembremos que en XIADA o lema está asociado á categoría gramatical. Por exemplo, o lema *actor* substantivo leva en fonte *volga_2004* mentres que no lema *actor* adxectivo figura *irindo_2004*, posto que no VOLGa só se recolle como substantivo.

proxecto⁸⁵, dispoñiamos dun método e protocolo de introdución de novas entradas no lexicón que nos facilitou enormemente a tarefa de actualización do lexicón e a inclusión masiva de novos elementos.

2.2.2.2. Categorias variables: modelos formais

2.2.2.2.1. A flexión de xénero e número

Polo que respecta ás clases nas que a maioría ou algún dos seus membros flexiona xénero e/ou número (Substantivo, Adxectivo, Indefinido, Relativo, Interrogativo-exclamativo, Posesivo, Numeral, Pronome, Demostrativo e Artigo), creamos grupos de derivación que facilitan esa flexión; nos grupos de derivación G1 a G28, G39 a G42 e mais o G44 e G45 recóllense terminacións para as cales se indica coa subetiqueta o xénero e o número correspondente, mentres que dende o grupo de derivación G29 ao G38 e no G46 só se achega para cada terminación a indicación de número. A numeración e establecemento dos distintos grupos derivacionais non responde á produtividade nin a ningunha orde lóxica, senón que estes foron establecidos aleatoriamente.

O elevado número de grupos de derivación creados anticipa que os factores tidos en conta para o seu establecemento non foron só as expresións dos morfemas flexivos de xénero e número, a todas luces insuficientes para acadar 45 combinacións. Esta cifra débese ás distintas expresións dos morfemas de xénero e número en combinación con alteracións gráficas de distinto tipo que sofren as formas ao flexionar. Por exemplo, pensemos nos substantivos *nen* e *rapaz*. A flexión destes lemas dá orixe ás formas seguintes:

	<i>Singular</i>	<i>Plural</i>
<i>Masculino</i>	nen rapaz	nenos rapaces
<i>Feminino</i>	nen rapaza	nenas rapazas

⁸⁵ Fco. Mario Barcala Rodríguez, compañeiro e amigo, é dende o ano 2000 o responsable informático do CORGA e XIADA. Sen o seu traballo moitos dos obxectivos lingüísticos perseguidos non serían posibles nin visibles.

En ambos os dous substantivos o morfema flexivo de xénero está expresado por *-a*: *nena*, *rapaza*. Na expresión do número, en cambio, o morfema flexivo presenta dous alomorfos: /s/ nos femininos e en *nenos* e /es/ en *rapaces*. Non obstante, non é á formación do plural unicamente á que debemos prestar atención para implementar as entradas correspondentes a *nenos* e *rapaz*. Temos que pensar que quen vai procesar a información é unha ferramenta informática e non un humano.

Por unha banda, para un humano a formación do feminino en *nenos* e *rapaz* é semellante, ambos se forman con *-a*. Pero para unha máquina os dous casos son moi diferentes, pois se os introducimos no mesmo grupo xerárase **nenoa* como feminino de *nenos*, porque os programas entenden literalmente a orde que se lles dá: “engadir *-a*”, sen ter en conta que en *nenos* primeiramente hai que eliminar o *-o* e despois engadir a expresión do morfema flexivo de xénero.

Por outra banda, na formación dos plurais *nenos* e *rapaces*, dende o punto de vista computacional, non só varía o alomorfo que se emprega nun e noutro caso, senón que tamén acada gran importancia a alternancia gráfica que se produce na raíz no lema *rapaz* entre o *z* do singular e o *c* do masculino plural, xa que está operando a regra grafemática segundo a cal para o fonema fricativo interdental xordo se usan dúas grafías:

c, seguido das vogais *e*, *i*: *rapaces*

z, seguido de *a*, *o*, *u* ou en posición final de palabra: *rapaz*, *rapaza*.

Por iso, as regras de flexión recollidas nos distintos manuais deben ser estudadas a fondo e matizadas. Por exemplo, nas *Normas* no apartado referente á formación do plural dise:

As rematadas en *-r* e *-z* engaden o morfema *-es* sobre o singular: *mar*, *mares*; *calor*, *calores*; *luz*, *luces*. [RAG/ILG, 2003: 75]

Para unha ferramenta informática, en troques, o plural das palabras rematadas en *-r* ten que tratarse de xeito diferente ao das rematadas en *-z*, porque no primeiro caso simplemente se engade *-es* mentres que no segundo a engádegas de *-es* para formar o plural masculino implica o cambio de *z* a *c*.

No establecemento dos múltiples grupos de derivación son, pois, determinantes non só as diferentes expresións para a flexión de xénero e número, senón tamén calquera modificación gráfica que poida ter lugar no proceso.

A *raíz* identifícase a partir da base léxica común ao paradigma flexivo e as terminacións están constituídas en realidade en numerosos casos polas expresións dos morfemas de xénero e número e mais a parte irregular do lexema tradicional. Por exemplo, por continuar co substantivo *rapaz*, en singular non presenta marcas de xénero masculino fronte ao feminino *rapaza*, termo marcado da oposición, que si presenta morfema flexivo de xénero. Ata aquí todo parece indicar que *rapaz* podería erixirse en *raíz*, en común denominador do grupo das catro formas flexionadas pertencentes ao lema *rapaz*. Mais mantense estable a raíz *rapaz* no plural? Si no feminino pero non no masculino, pois neste último debido ás leis grafemáticas o *-z* final transfórmase en *c*. Non nos vale polo tanto *rapaz* como *raíz* polo que debemos preguntarnos que parte da palabra é, entón, común ás catro formas flexionadas. A resposta, claro, é obvia: *rapa*. Deste xeito *rapa* fórmase como *raíz* por presentarse nel o segmento común a *rapaz*, *rapaza*, *rapaces* e *rapazas*.

Como se deciden as terminacións que deben configurar un grupo? Prescindindo da raíz e fixándonos nas partes variables das formas comúns ao mesmo paradigma. Continuando co lema *rapaz*, se prescindimos da raíz *rapa* que acabamos de determinar, quédanos *z* para expresar o masculino singular, *za* para o feminino singular e *ces* e *zas* para os plurais masculino e feminino, respectivamente, co que xa temos establecido o contido dun dos grupos de derivación que operan no lexicón de XIADA, o G22:

Grupo de derivación: G22		
Representante de grupo: <i>rapaz</i>		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i> ⁸⁶
z	masculino singular	<i>rapa-z</i>
ces	masculino plural	<i>rapa-ces</i>
za	feminino singular	<i>rapa-za</i>
zas	feminino plural	<i>rapa-zas</i>

⁸⁶ As formas que de aquí en adiante se recollen baixo esta cabeceira nas táboas exemplificadoras dos grupos de derivación non aparecen en ningún caso no lexicón. Tampouco o guión que separa visualmente en cada forma a raíz da terminación. A súa presenza débese a unha soa razón: facilitar a lectura e comprensión.

Este grupo, o G22, caracteriza no noso lexicón 20 formas pertencentes a 5 lemas diferentes, 4 substantivos (*rapaz, xuíz, aprendiz, andaluz*) e 1 adxectivo (*andaluz*).

Tal como acabamos de ver, os criterios que rexen o establecemento de raíces e terminacións para a implementación dos elementos pertencentes ás clases flexivas non teñen nada que ver coa segmentación tradicional, e mesmo non son defendibles dende un punto de vista lingüístico. Con todo, o noso labor non consiste en realizar rigorosas análises morfolóxicas en constituíntes, senón en describir a gramática formal do galego de xeito tal que a información que se proporcione sexa real, completa e compatible cunha aplicación informática. Priman neste caso, xa que logo, as necesidades computacionais sobre a rigorosidade lingüística.

Non todos os grupos de derivación establecidos presentan niveis semellantes de produtividade; uns revélanse como o recurso produtivo para a formación de pares léxicos opostos polo xénero e/ou o número, mentres que outros reproducen casos fixados no devir histórico da lingua (*galo/galiña*). No sistema de XIADA, por coherencia pero tamén por deixar unha porta aberta á posible inclusión de novos elementos, procuramos reflectir en grupos todas as oposicións xenéricas ou numéricas expresadas pola mesma base léxica. Vexámoslos.

A regra xeral de formación do feminino, segundo a cal para os masculinos rematados en *-o* ou *-e* se elimina esa vogal e se engade *-a* (Álvarez, Regueira e Monteagudo, 1986: 62-63), á que lle corresponde na formación do plural engadir *-s*, require para a súa implementación no sistema de XIADA a creación de dez grupos de derivación distintos: G1, G4, G9, G13, G15, G16, G19, G20, G24 e G25.

Por unha banda, temos que separar en dous grupos distintos os masculinos que rematan en *-o* (G1) dos que o fan en *-e* (G13), pois en ambos os casos a vogal final non pode formar parte da raíz por ter que eliminala para construír o feminino:

Grupo: G1		
Representante: neno		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
o	masculino singular	<i>nen-o</i>
os	masculino plural	<i>nen-os</i>
a	feminino singular	<i>nen-a</i>
as	feminino plural	<i>nen-as</i>

Grupo: G13		
Representante: este		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
e	masculino singular	<i>est-e</i>
es	masculino plural	<i>est-es</i>
a	feminino singular	<i>est-a</i>
as	feminino plural	<i>est-as</i>

O G1 é con moito o máis produtivo dos grupos de derivación creados, pois a el remítense as entradas correspondentes a 11 156 lemas ou se se prefire, o G1 permite caracterizar morfoloxicamente e mais xerar 44 624 formas, sen incluír nesta cantidade todos os participios regulares que se envían ao G1, porque figuran no interior dos distintos grupos verbais, motivo polo que non deben contabilizarse aquí e si nas estatísticas dos grupos verbais. Ademais, o G1 é o único que acolle elementos de case todas as clases de palabras nas que se produce flexión de xénero e número, faltándolle unicamente representar o pronome persoal.

11 156 lemas	Categoría	Exemplo
9302	Adxectivo	<i>amarelo</i>
1744	Substantivo	<i>año</i>
55	Verbo	<i>aberto</i>
30	Numeral	<i>primeiro</i>
19	Indefinido	<i>mesmo</i>
2	Relativo	<i>cuxo</i>
2	Interrogativo	<i>canto</i>
2	Posesivo	<i>noso</i>

A pregunta que subxace é por que non se envía a este grupo ningunha das formas do paradigma pronominal, no que nos consta que si hai flexión de xénero e número coma a representada no G1? A resposta é sinxela. Dos tres alomorfos do pronome átono de terceira acusativo (*o, lo, no*) só o primeiro, *o, a, os, as* pode aparecer illado, ou sexa sen concorrer cunha forma verbal ou integrado nunha contracción. Na base de datos léxica de XIADA soamente introducimos como entradas aquelas cuxas formas son susceptibles de aparecer escritas entre dous espazos en branco; as formas que, por atoparse en distribución complementaria, precisan de amalgamárense a outros elementos para poder expresarse non figuran no núcleo do lexicón de XIADA, senón

que son elementos externos a el e, como veremos máis adiante, reciben un tratamento específico, diverxente do do vocabulario común.

Isto explicaría por que *lo, la, los, las* (alomorfos do pronome ou do artigo) e *no, na, nos, nas* (alomorfo do pronome) non se remiten ao G1. Mais que ocorre coa primeira forma, co artigo ou pronome *o, a, os, as*? É certo que seguen o modelo flexional establecido para o G1, pero nas formas singulares é imposible discriminar na expresión unha raíz e unha terminación, pois o masculino e feminino singulares só presentan un carácter. Por este motivo as entradas *o* e mais *a*, cuxo lema é *o*, coas dobres subetiquetas pertencentes ás clases Pronome e Artigo remítense a un grupo de derivación no que só opera o número: o G29.

O G13, pola súa banda, contén membros das clases Substantivo, Adxectivo e Demostrativo e, incluso sen ser dos menos frutíferos, recibe poucos elementos, pois só caracteriza as formas correspondentes a 26 lemas:

26 lemas	Categoría	Exemplo
20	Substantivo	<i>cliente</i>
4	Adxectivo	<i>mestre</i>
2	Demostrativo	<i>ese</i>

Por outra parte e continuando coa formación do feminino, os masculinos rematados en *-o* átono, deixando xa de lado o G1, debemos desagregalos aínda en tres grupos máis: G4, G16 e G25.

O G4, cuxo representante é LANCARAO, xorde por alterarse a formación do feminino en *-a*, xa que por crase dos dous *-aa* o feminino transfórmase en *-á*:

Grupo: G4		
Representante: lancrao		
terminación	subetiqueta	exemplo
ao	masculino singular	<i>lancar-ao</i>
aos	masculino plural	<i>lancar-aos</i>
á	feminino singular	<i>lancar-á</i>
ás	feminino plural	<i>lancar-ás</i>

E, finalmente, o G16 e o G25 responden á formación do feminino por medio dun morfema derivativo, distinto en cada un dos grupos, sobre a mesma base léxica ca o masculino correspondente:

Grupo: G16		
Representante: galo		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
o	masculino singular	<i>gal-o</i>
os	masculino plural	<i>gal-os</i>
iña	feminino singular	<i>gal-iña</i>
iñas	feminino plural	<i>gal-iñas</i>

Grupo: G25		
Representante: vampiro		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
o	masculino singular	<i>vampir-o</i>
os	masculino plural	<i>vampir-os</i>
esa	feminino singular	<i>vampir-esa</i>
esas	feminino plural	<i>vampir-esas</i>

O G4 permite a caracterización e xeración de 104 formas que se remiten a 26 lemas distintos, 12 deles asociados á clase Substantivo (*limiao*) e 14 á de Adxectivo (*mao*), mentres que o G16 e o G25 só caracterizan as 4 formas que se xeran en cada un dos grupos respectivos, pertencentes ao paradigma de *galo* e *vampiro*.

Cando o masculino remata en *-e*, á parte do G13 xa visto, debemos considerar tres grupos máis, novamente debido á formación do feminino mediante un morfema derivativo diferente en cada un dos tres casos: o G9, mediante o que caracterizamos substantivos coma *alcalde*, *duque*, *tigre*, etc.; o G19, co cal definimos *heroe* e *superheroe* e o G24, que só empregamos para as 4 formas do lema *sacerdote*.

Grupo: G9			Grupo: G19			Grupo: G24		
Representante: alcalde			Representante: heroe			Representante: sacerdote		
<i>terminación</i>	<i>subeti.</i>	<i>exemplo</i>	<i>terminación</i>	<i>subeti.</i>	<i>exemplo</i>	<i>terminación</i>	<i>subeti.</i>	<i>exemplo</i>
e	masculino singular	<i>alcald-e</i>	e	masculino singular	<i>hero-e</i>	e	masculino singular	<i>sacerdot-e</i>
es	masculino plural	<i>alcald-es</i>	es	masculino plural	<i>hero-es</i>	es	masculino plural	<i>sacerdot-es</i>
esa	feminino singular	<i>alcald-esa</i>	ína	feminino singular	<i>hero-ína</i>	isa	feminino singular	<i>sacerdot-isa</i>
esas	feminino plural	<i>alcald-esas</i>	ínas	feminino plural	<i>hero-ínas</i>	isas	feminino plural	<i>sacerdot-isas</i>

Mais non terminan aínda aí as peculiaridades flexivas dos masculinos rematados en vogal, xa que segundo Álvarez, Regueira e Monteagudo (1986: 63) os rematados en vogal tónica engaden *-a* sen alterar a base: *avó/avoa*, coa salvidade de *xudeu/xudía* que

constitúe unha excepción a esta regra. Mentres tanto nas *Normas* (RAG/ILG, 2003: 81) indícase que “*os substantivos rematados en vogal tónica tenden a engadir -a, en tanto que os adxectivos tenden a ser invariables: un avó somalí / unha avoa somalí. Con todo, algúns substantivos son invariables: unha chimpancé adulta; entre os adxectivos, nu, cru e recrú, só e mao ‘ruín’ forman os femininos núa, crúa e recrúa, soa e má: o peixe cru / a carne crúa, un só día / unha casa soa, o ollo mao / unha má ocasión. Presentan características morfolóxicas especiais grou / grúa, xudeu / xudía e sandeu / sandía.*”

Vaiamos por partes, pois a regra enunciada no parágrafo anterior, á hora de aplicala a un sistema informático, presenta matices interesantes. Por un lado, a non alteración da base en *avó/avoa* seguindo a *Gramática galega* (Álvarez, Regueira e Monteagudo, 1986: 63) non é exactamente así para unha aplicación computacional. A formación do feminino en *avó* provoca que a palabra deixe de ser aguda e se transforme en grave, co que as regras de acentuación subxacentes varían do masculino ao feminino e implican un cambio na base léxica, pequeno pero moi importante a efectos computacionais. Este é o motivo de que no lexicón de XIADA creamos dous grupos para os rematados en *-o* tónico: o G23 e o G40. No G23, cuxo representante de grupo é AVÓ, o masculino singular e plural acentúase graficamente, mentres que no G40, cuxo representante é BO, por ser monosílabo o masculino, non se acentúan ningunha das formas resultantes:

Grupo: G23		
Representante: avó		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ó	masculino singular	<i>av-ó</i>
ós	masculino plural	<i>av-ós</i>
oa	feminino singular	<i>av-oa</i>
oas	feminino plural	<i>av-oas</i>

Grupo: G40		
Representante: bo		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplos</i>
	masculino singular	<i>bo, recrú</i>
s	masculino plural	<i>bo-s, recrú-s</i>
a	feminino singular	<i>bo-a, recrú-a</i>
as	feminino plural	<i>bo-as, recrú-as</i>

Ningún destes dous grupos podemos consideralo produtivo, xa que o G23 caracteriza 4 lemas —os substantivos *avó, bisavó, tataravó* e o adxectivo *só*—, mentres que o G40 unicamente describe as 8 formas adxectivas correspondentes aos lemas *bo* e *recrú*.

Paralelamente, para os masculinos rematados en *-u* tónico, tivemos que establecer un novo grupo de derivación diferenciando así entre os monosílabos e os polisílabos, xa que aínda que a efectos de acentuación coinciden os femininos singulares e plurais en ambos os casos, non ocorre así cos masculinos: *cru(s)* fronte a *recrú(s)*. O G44 facilita a caracterización e xeración dos monosílabos *cru* e *nu*, mentres que o polisílabo *recrú* se integra no G40 por coincidiren *bo* e *recrú* nas terminacións e valores caracterizadores destas.

Grupo: G44		
Representante: <i>cru</i>		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
u	masculino singular	<i>cr-u</i>
us	masculino plural	<i>cr-us</i>
úa	feminino singular	<i>cr-úa</i>
úas	feminino plural	<i>cr-úas</i>

Sorprende nas *Normas* (RAG/ILG, 1995: 82; 2003: 81) a inclusión do adxectivo *mao* na epígrafe correspondente á acentuación dos nomes rematados en vogal ou ditongo tónicos, posto que nin remata en vogal tónica nin incrementa o morfema flexivo de xénero sen previamente prescindir do denominado en ocasións morfo de xénero masculino (Álvarez e Xove, 2002).

Polo que se refire ás palabras agudas rematadas en ditongo que flexionan o xénero, ademais de ser o seu número moi escaso, parece que a formación do feminino é totalmente irregular: *rei/raiña*, *xudeu/xudía*, *sandeu/sandía* e *grou/grúa*. Fóra os posesivos *meu*, *teu* e *seu*, que posúen un feminino *miña*, *túa* e *súa* que tampouco se axustan a ningún dos grupos anteriores, non coñecemos ningunha outra palabra que remate en ditongo e flexione o xénero mediante morfema flexivo. A consulta á versión 2.2 do léxico descargable de XIADA (Rojo, Domínguez e Barcala, 2009) así como ao *Diccionario inverso da lingua galega* (ILG/RAG, 1989) ou ao VOLGa en liña⁸⁷ non amosou resultados diferentes.

⁸⁷ Dispoñible en <http://www.realacademiagalega.org/volga> [consultado o 19/07/2010]. A consulta realizada, calquera expresión rematada nas vogais *-i* ou *-u*, non é posible levala a cabo na edición en papel polo que recorreremos á versión en liña, aínda que desta última non podemos especificar de que versión se trata ou en que data foi actualizada, posto que esa información non aparece na páxina web.

En resumidas contas, a formación do feminino para os rematados en ditongo obríganos a crear catro grupos para cubrir outras tantas irregularidades: o G15, cuxo representante é XUDEU, caracteriza 3 lemas, 2 asociados á clase Adxectivo (*xudeu* e *sandeu*) e 1 á clase Substantivo (*xudeu*); o G17, tipo REI, caracteriza o paradigma flexivo dos lemas *rei* e *vicerrei*; o G20 está representado e configurado unicamente polo lema GROU; e, no último lugar desta serie, o G45, cuxo representante é o posesivo TEU, acolle a caracterización dos posesivos distributivos *cadanseu* e *cadaseu* –variante non normativa recollida no VOLGa (González e Santamarina, 2004) e empregada nos textos do CORGA– así como a dos posesivos *teu* e mais *seu*. O posesivo de primeira persoa singular, *meu*, non se pode incluír no G45 por diverxer o feminino, mais nesta ocasión, por preguiza fundamentalmente, non creamos aínda un grupo novo ao que incorporar a flexión de *meu*, xa que nos era máis cómodo e rápido introducilo no lexicón mediante dúas entradas: unha pertencente ás formas masculinas e outra correspondente ás femininas, remitindo ambas ao grupo G29 onde se flexiona o número.

Grupo: G15		
Representante: xudeu		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
eu	masculino singular	<i>xud-eu</i>
eus	masculino plural	<i>xud-eus</i>
ía	feminino singular	<i>xud-ía</i>
ías	feminino plural	<i>xud-ías</i>

Grupo: G17		
Representante: rei		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ei	masculino singular	<i>r-ei</i>
eis	masculino plural	<i>r-eis</i>
aíña	feminino singular	<i>r-aíña</i>
aíñas	feminino plural	<i>r-aíñas</i>

Grupo: G20		
Representante: grou		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ou	masculino singular	<i>gr-ou</i>
ous	masculino plural	<i>gr-ous</i>
úa	feminino singular	<i>gr-úa</i>
úas	feminino plural	<i>gr-úas</i>

Grupo: G45		
Representante: teu		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
eu	masculino singular	<i>t-eu</i>
eus	masculino plural	<i>t-eus</i>
úa	feminino singular	<i>t-úa</i>
úas	feminino plural	<i>t-úas</i>

Para concluír a descrición dos grupos de derivación de XIADA cuxo masculino singular remata en vogal só nos queda dar conta do G11, tipo PROFETA, non incluído

no mesmo apartado ca SACERDOTE (G24) ou VAMPIRO (G25), a pesar de tamén formar o feminino mediante un morfema derivativo, por estar tratando naquel momento a formación do feminino a partir dos masculinos rematados en *-o* ou *-e*. Co G11 cóbrese a caracterización morfolóxica das formas pertencentes a 3 lemas substantivos: *profeta*, *papa* e *poeta*.

Grupo: G11		
Representante: profeta		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
a	masculino singular	<i>profet-a</i>
as	masculino plural	<i>profet-as</i>
isa	feminino singular	<i>profet-isa</i>
isas	feminino plural	<i>profet-isas</i>

No referente aos grupos de derivación orixinados pola flexión xenérica e/ou numérica dos masculinos rematados en consoante, comezaremos cos acabados en *-r*, *-s* ou calquera outra consoante para os cales o feminino se forma engadindo *-a* e que no plural acrecentan *-s* para o feminino e *-es* para o masculino. Cumpre esta característica o grupo de derivación G6, cuxo representante é AQUEL, e que despois do G1 resulta ser o máis produtivo con 1624 lemas, distribuídos en 854 adxectivos (*abrumador*), 768 substantivos (*adestrador*), 1 demostrativo (*aquel*) e 1 pronome (*el*).

Grupo: G6		
Representante: aquel		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
	masculino singular	<i>aquel</i>
es	masculino plural	<i>aquel-es</i>
a	feminino singular	<i>aquel-a</i>
as	feminino plural	<i>aquel-as</i>

Tamén xuntan *-es* como expresión do plural masculino os grupos G12 (representante XOGRAR) e G18 (representante TSAR), mais a formación do feminino realízase neles mediante un sufixo derivativo: *xograresa*, *tsarina*. O G12 acolle, ademais da flexión de *xograr*, a de *prior*, mentres que o G18 alberga a de *tsar* e mais a da variante non normativa actualmente *zar*.

Grupo: G12		
Representante: xograr		
terminación	subetiqueta	exemplo
	masculino singular	<i>xograr</i>
es	masculino plural	<i>xograr-es</i>
esa	feminino singular	<i>xograr-esa</i>
esas	feminino plural	<i>xograr-esas</i>

Grupo: G18		
Representante: tsar		
terminación	subetiqueta	exemplo
	masculino singular	<i>tsar</i>
es	masculino plural	<i>tsar-es</i>
ina	feminino singular	<i>tsar-ina</i>
inas	feminino plural	<i>tsar-inas</i>

Non cumpre tampouco cun feminino en *-a* o modelo G27, tipo EMPERADOR, nin o G28, tipo ACTOR. Ademais, a pesar das similitudes na formación do feminino, non podemos integralos no mesmo grupo, pois no primeiro caso prodúcese a substitución do fonema oclusivo dental sonoro, presente no masculino, polo oclusivo dental xordo do feminino: *emperador* / *emperatriz*, mentres que en *actor* mantense sempre a consoante dental xorda, co que teñen que diferir necesariamente, desde un punto de vista computacional, a raíz e o contido dos grupos caracterizadores de *emperador* e *actor*.

Grupo: G27		
Representante: emperador		
terminación	subetiqueta	exemplo
dor	masculino singular	<i>empera-dor</i>
dores	masculino plural	<i>empera-dores</i>
triz	feminino singular	<i>empera-triz</i>
trices	feminino plural	<i>empera-trices</i>

Grupo: G28		
Representante: actor		
terminación	subetiqueta	exemplo
or	masculino singular	<i>act-or</i>
ores	masculino plural	<i>act-ores</i>
riz	feminino singular	<i>act-riz</i>
rices	feminino plural	<i>act-rices</i>

O G27 só describe no noso lexicón as formas flexionadas de 2 lemas, o substantivo *emperador* e o adxectivo *osculador*, entre tanto o G28 caracteriza 6 lemas, o substantivo e adxectivo *actor* e os adxectivos *electromotor*, *magnetomotor*, *psicomotor* e *vasomotor*.

Despois do G1 e o G6, o G5, cuxo representante é INGLÉS, eríxese no terceiro grupo máis frutífero, pois a el remítese a caracterización e xeración de 355 lemas, dos cales 172 pertencen á clase Substantivo (*brionés*) e 183 á de Adxectivo (*canchés*). Este modelo caracterízase por presentar no seu paradigma un masculino singular acentuado

fronte ás demais formas: *inglés /inglesa(s), ingleses*. Nel, ademais, a formación do plural difire do masculino –engade *-es*– ao feminino –suma *-s*–:

Grupo: G5		
Representante: inglés		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
és	masculino singular	<i>ingl-és</i>
eses	masculino plural	<i>ingl-eses</i>
esa	feminino singular	<i>ingl-esa</i>
esas	feminino plural	<i>ingl-esas</i>

Para os masculinos rematados en *-n*, temos que diferenciar entre os que acaban en *-án*, os que finalizan en *-ín* e os que terminan en *-ón*.

Unicamente os que conclúen o masculino en *-ín* presentan homoxeneidade na flexión do xénero e o número, pois para o feminino engaden *-a* e para os respectivos plurais *-s*. Malia que o seu comportamento parece, polo que acabamos de indicar, idéntico aos que rematan en vogal tónica (tipo AVÓ, G23), as terminacións entre ambos diverxen. No G10, cuxo representante de grupo é BAILARÍN, debido á diversa acentuación das formas do paradigma e ao remate da palabra en *-n*, a terminación masculina singular está constituída polo segmento *-ín* mentres que a feminina singular é *-ina*. O grupo G10 acolle a flexión de 15 lemas distribuídos en 8 substantivos (*benxamín*) e 7 adxectivos (*cantarín*):

Grupo: G10		
Representante: bailarín		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ín	masculino singular	<i>bailar-ín</i>
íns	masculino plural	<i>bailar-íns</i>
ina	feminino singular	<i>bailar-ina</i>
inas	feminino plural	<i>bailar-inas</i>

Polo que se refire aos masculinos rematados en *-án*, presentan dous resultados diferentes na formación normativa do feminino: *-á* (G2, tipo CATALÁN) e *-ana* (G3, tipo PILLABÁN).

Adiantaremos aquí o contido dos dous grupos porque é nesta epígrafe onde comentamos os grupos derivacionais creados para facilitar a introdución do léxico no

noso sistema, mais imos postergar a súa explicación ao apartado relativo ás repercusións que provocou no lexicón de XIADA a edición das *Normas* do 2003 (RAG/ILG, 2003) para non repetirmos pero, sobre todo, porque a aparición desta edición provocou cambios nas terminacións do G2 e G3 e nas remisións de certos vocábulos a estes grupos.

Grupo: G2			Grupo: G3		
Representante: catal-án			Representante: pillab-án		
terminación	subetiqueta	normativa2003	terminación	subetiqueta	normativa2003
án	masculino singular	si	án	masculino singular	si
áns	masculino plural	si	áns	masculino plural	si
á	feminino singular	si	ana	feminino singular	si
ás	feminino plural	si	anas	feminino plural	si
ana	feminino singular	non	á	feminino singular	non
anas	feminino plural	non	ás	feminino plural	non

Coa terminación feminina normativa *-ana*, grupo G3 polo tanto, figuran en XIADA 123 lemas repartidos categorialmente en 60 substantivos (*chalán*) e 63 adxectivos (*brután*), mentres que coa terminación *-á* (G2), tamén feminina e normativa, rexístranse 120 lemas repartidos en 60 substantivos (*capitán*) e 60 adxectivos (*alazán*). A produtividade, como vemos, é parella.

Os masculinos rematados en *-ón* con flexión de xénero e número superan lixeiramente os finalizados en *-án*, pois fronte aos 243 lemas totais vistos para *-án*, os de *-ón* suman 271 distribuídos do seguinte xeito:

Grupo	Representante	Total de lemas	Substantivos	Adxectivos
G7	LAMBÓN	217	101	116
G8	CAMPIÓN	51	27	24
G14	CABRÓN	3	2	1

Cos masculinos rematados en *-án* (G2, G3) e mais os non pexorativos acabados en *-ón* (G8) que flexionan o xénero, as *Normas* do ano 2003 (RAG/ILG, 2003) creáronnos un pequeno problema no concerne á clasificación e caracterización de certos elementos que se remitían a estes grupos, posto que na nova edición, nalgúns casos, modificouse lixeiramente a anterior formación do feminino. Non imos tratar aquí, como xa avisamos, o contido dos grupos G2 e G3 nin tampouco o do G7 e G8 porque

estes aparecerán explicados nun apartado posterior en relación cos cambios introducidos pola última edición das *Normas* (RAG/ILG, 2003), non obstante, é imprescindible para poder entender o problema que trataremos a continuación recoller, aínda que de xeito sucinto, as desinencias e subetiquetas que conforman o G7 e o G8:

Grupo: G7			Grupo: G8		
Representante: lamb-ón			Representante: campi-ón		
<i>terminación</i>	<i>subetiqueta</i>	<i>normativa2003</i>	<i>terminación</i>	<i>subetiqueta</i>	<i>normativa2003</i>
ón	masculino singular	si	ón	masculino singular	si
óns	masculino plural	si	óns	masculino plural	si
ona	feminino singular	si	oa	feminino singular	si
onas	feminino plural	si	oas	feminino plural	si
			ona	feminino singular	non
			onas	feminino plural	non

O G14, continuando coa descrición dos grupos, xorde no noso sistema para cubrir os femininos irregulares de CABRÓN –representante do grupo– e LADRÓN: *cabra* e *ladra*. Estes non teñen cabida no G8, tipo CAMPIÓN, ao lado dos femininos regulares en *-oa*, nin no G7, tipo LAMBÓN, cos femininos en *-ona*, xa que son os únicos lemas en *-ón* que posúen dobre feminino: *cabrona/cabra* e *ladroa/ladra*. Por suposto, non podemos esquecernos de que *ladrón* posúe aínda a maiores o feminino *ladrona*, non normativo actualmente pero si ata o 2003.

A pregunta que podería formularse é por que un grupo propio e non a adaptación das terminacións no G7 (cabrón) e G8 (ladrón) para incluír a maiores os femininos *ladra* e *cabra*. A resposta ten que ver coa xeración de formas, proceso complementario ao da caracterización morfolóxica. É certo que aínda que introduzamos unha terminación *-a* no G7 e G8, se nos documentos non aparecen, tampouco nos crearía excesivos problemas para a identificación e caracterización dos *tokens* na etiquetaxe de corpus. O problema xorde á hora de xerar as formas procedentes do lexicón, porque ou ben restrinximos computacionalmente a terminación *-a* de xeito que só sexa posible para os lemas *ladrón* e *cabrón* ou ben permitimos que se empregue con todas as entradas que se remiten a eses grupos. Esta última opción non é aceptable porque as formas ás que daría lugar non existen na nosa lingua (de *lambón*, **lamba*; de *campión*, **campia*, etc.). Tampouco nos parece adecuada a primeira opción, a da restrición para

uns determinados lemas, porque constituiría unha incoherencia no sistema de XIADA xa que se erixiría no único grupo no que algunha das súas terminacións condiciona o lema, cando precisamente esta condición é no noso lexicón a que provoca o establecemento dos distintos grupos. De aí que o G14 teña entidade propia e presente o seguinte contido:

Grupo: G14		
Representante: <i>cabrón</i>		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ón	masculino singular	<i>cabr-ón</i>
óns	masculino plural	<i>cabr-óns</i>
a	feminino singular	<i>cabr-a</i>
as	feminino plural	<i>cabr-as</i>

Temos que recoñecer que a implementación do representante de grupo do G14, *cabrón*, causounos máis dun problema, aínda que tamén é verdade que no proceso aprendemos unhas cantas cousas. Por unha banda, no relativo á identificación de lemas, estamos ante dous lemas distintos pertencentes á mesma clase de palabra: Substantivo. Tanto na súa acepción de animal (fem. *cabra*) como na de persoa que obra con mala intención (fem. *cabrona*), *cabrón* eríxese en lema de dúas entradas distintas en XIADA, co que en consecuencia temos no noso sistema dous lemas *cabrón*. Non debe estrañarnos esta situación porque non é única; ocorre tamén por exemplo co lema *galo* que presenta dúas entradas en XIADA, unha, a de animal doméstico cun feminino *galiña* e outra, a do xentilicio cun feminino *gala*.

Por outra banda, no concernente á fonte do lema, atopámonos con que o VOLGa (González e Santamarina, 2004), empregado no noso sistema como fonte básica do léxico común, non inclúe como entrada a posibilidade de *cabrón cabrona* mentres que si recolle *cabrón* como substantivo masculino desligado totalmente do substantivo feminino *cabra*, que tamén inclúe.

A non presenza no VOLGa de “*cabrón a s f: cabrona*” a semellanza de “*lambón a s f: lambona*” levounos a pensar que foi un *lapsus* ou que nesta acepción a entrada non era normativa, polo que consultamos a cuestión cun dous autores. O profesor González González moi amablemente, coma sempre, confirmounos que si era normativa e que en posteriores edicións se incluiría. Con isto evitamos o problema de caracterizar como

non normativas as formas *cabrona* e *cabronas* pertencentes paradoxalmente a un lema normativo. Como é isto posible? A explicación ten a súa orixe na non introdución de información semántica para os distintos lemas. Se lembramos, no noso sistema o lema está asociado á categoría e non a acepcións, de xeito que se difiren as categorías necesariamente estamos ante lemas distintos. O gran problema prodúcese cando coincide lema e categoría pero non a fonte do lema ou a normatividade da forma. Nestes casos nós non podemos marcar directamente no sistema unha fonte distinta para un lema que xa existe e cuxa subetiqueta coincide, aínda que se remitan a grupos distintos. Tampouco podemos marcar como non normativo un lema que coincide con outro lema xa existente se ademais coinciden na categoría, por moito que as formas que se poidan orixinar sexan distintas nun caso e noutro.

Continuando co lema *cabrón*, consultamos outras obras lexicográficas, porque na medida do posible desexamos que a información que introducimos en XIADA estea documentada.

No DRAG (García e González, 1997) baixo a entrada *cabrón*, definida categorialmente como substantivo masculino, recóllense as seguintes acepcións:

1. Macho da cabra.
2. Insulto dirixido a unha persoa que fai algo con mala intención para enfastiar a outra.
3. Marido ó que a muller lle é infiel, particularmente cando o consente.

Engádesse ademais unha observación na que se indica que “*as acepcións 2 e 3, de uso vulgar, fan que a palabra sexa tabú mesmo na primeira acepción*”. Como se observa, non se recolle *cabrona* nin se achega ningunha información sobre a forma feminina correspondente, a pesar de na acepción 2 ser unha ‘persoa’ o referente e non o masculino ‘home’.

No Xerais do ano 2000 na voz *CABRÓN* non se informa sobre o seu feminino e as acepcións que recolle son as que seguen:

CABRÓN s.m. 1. Macho da cabra; bode, castrón. // *adx.* e *s.* 2. *fig.* Home que consente no adulterio do seu cónxuxe; cornudo. // *s.* 3. Persoa maliciosa ou malintencionada.

No do ano 2009 introducen pequenas variacións nas definicións, mais séguese sen facer ningunha aclaración sobre o feminino:

CABRÓN [ka'βron] *s.m.* 1. Macho da cabra apto para a procreación; bode, castrón, chibo. // *adx.* e *s.* 2. *fig.* e *fam.* Home que consente o adulterio do outro cónxuxe; castrón. // *s.* 3. Persoa maliciosa ou malintencionada. **Facer de c. [algo]** *loc. v.*, facelo con malicia.

Pola contra, tanto na edición do 2000 como na do 2009 de Xerais si atopamos xa na entrada a formación feminina de voces como *LAMBÓN -ONA* ou *LADRÓN -OA*, xunto cunha observación nesta última voz sobre a constatación de *ladra* como outra forma de feminino.

No Ir Indo (Ledo Cabido, 2004), así como na súa versión en liña⁸⁸, xa se introduce o feminino na entrada da voz: *cabrón -na*, mais ese feminino parece ser válido só para a segunda acepción, pois tanto a primeira como a terceira son categorialmente substantivos masculinos:

cabrón -na 1 *s m* Macho da cabra. 2 *s* Persoa que obra con mala intención. *Tm adx.* 3 *s m* Home que consente o adulterio. *Tm adx.*

Á parte da curiosidade que nos domina por saber como se denomina a muller que consente o adulterio do cónxuxe, porque por máis que preguntamos non obtivemos resposta, o que se evidencia da consulta realizada en distintas obras lexicográficas actuais é que non hai consenso á hora de tratar os femininos irregulares, e en ocasións tampouco os complementarios. Non é a nosa intención incidir sobre o sexismo da lingua, tema que está moi de actualidade e que incluso foi obxecto dunhas xornadas (Mosquera, 2010), mais si nos interesa o feito de que a xeneralización da alternancia xenérica nos nomes con referencias persoais non sempre aparece ben resolta, como podemos comprobar na voz *cabrón*.

Hermida (2010), nun interesante artigo sobre o sexismo nos dicionarios galegos, pon de manifesto que no lema *cabrón* non se resolven axeitadamente os problemas derivados dunha correcta adscrición dos significados coa categoría correspondente e que “*a fórmula de presentación que nos parece correcta e respectuosa cos usos que existen na fala é a seguinte:*

CABRÓN¹ *s.m.* 1. Macho da cabra apto para a procreación; bode, castrón, chibo. // *adx.* e *s.m.* 2. *fig.* e *fam.* Home que consente o adulterio do outro cónxuxe; castrón.

⁸⁸ Primeiramente libre e gratuita, e logo accesible mediante subscrición en <http://www.digalego.com> [consultado o 22/07/2010].

CABRÓN² -NA s. Persoa maliciosa ou malintencionada. **Facer de c. [algo] loc. v.**, facelo con malicia.”

Con esta representación resolveríamos parte do problema co que nos tivemos que afrontar, pero só parte porque seguímonos preguntando por que non se relacionan dalgún xeito nos dicionarios os femininos irregulares cos masculinos correspondentes. É dicir, sabemos que *cabra* é o feminino de *cabrón*, non obstante, por moito que *cabra* sexa o xenérico para denominar esa especie, por que nin en *cabra* nin en *cabrón* se relacionan ambas as voces? Iso implica que tampouco nós as deberíamos ligar no lexicón? Por que no DRAG (1997), VOLGa (2004) ou no Xerais (2000 e 2009) aparecen entradas diferenciadas para *cabrón*, *cabra*, *rei*, *raíña*, *grou*, *grúa*, etc. e no Ir Indo en liña están ligadas? Por que no VOLGa (2004) en *cabrón*, *cabra* ou *carneiro* e *ovella* non hai nin en xénero complementario nin en notas nada á parte do feminino complementario *carneira* e en *grou* aparece en nota “paxaro f grúa” e en *grúa* aparece “paxaro m grou”?

Retomando a descrición dos grupos de derivación en XIADA, para os masculinos rematados en *-ón* individualizamos no G26 a flexión do lema BARÓN, motivado pola engádega dun morfema derivativo na formación do feminino:

Grupo: G26		
Representante: barón		
terminación	subetiqueta	exemplo
ón	masculino singular	<i>bar-ón</i>
óns	masculino plural	<i>bar-óns</i>
onesa	feminino singular	<i>bar-onesa</i>
onesas	feminino plural	<i>bar-onesas</i>

Co G21, cuxo representante é ESPAÑOL, posibilitamos a caracterización e xeración das formas flexionadas dos substantivos *español* e *colexial* e mais o adxectivo *español*. A súa creación responde á necesidade de dar cabida á presenza do *-l* nos singulares e mais no plural feminino e á súa ausencia no plural masculino:

Grupo: G21		
Representante: español		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
l	masculino singular	<i>español-l</i>
is	masculino plural	<i>español-is</i>
la	feminino singular	<i>español-la</i>
las	feminino plural	<i>español-las</i>

O adxectivo *colexial*, a pesar de non diferenciarse xenericamente do substantivo *colexial* no VOLGa⁸⁹ ou na versión en liña do Ir Indo, tal como se recolle nunha observación na entrada *colexial-ala* do DRAG (1997: 276), é invariable en canto ao xénero, non presenta un feminino *colexiala*, polo que non pode integrarse no mesmo grupo.

O G39, cuxo representante é ALGÚN, permite a identificación e xeración do paradigma dos indefinidos normativos *algún* e *ningún* así como do non normativo *nengún*:

Grupo: G39		
Representante: algún		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
ún	masculino singular	<i>alg-ún</i>
úns	masculino plural	<i>alg-úns</i>
unha	feminino singular	<i>alg-unha</i>
unhas	feminino plural	<i>alg-unhas</i>

O artigo indefinido ou o indefinido *un*, remitidos ao grupo G41, non poden integrarse no mesmo modelo ca o indefinido *algún* debido ás diferenzas que establece a acentuación:

Grupo: G41		
Representante: un		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
	masculino singular	<i>un</i>
s	masculino plural	<i>un-s</i>
ha	feminino singular	<i>un-ha</i>
has	feminino plural	<i>un-has</i>

⁸⁹ Nas entradas clasificadas categorialmente con "a" (=adxectivo) ou "a s" (=adxectivo e substantivo) rematadas en *-al*, o único caso para o que se achega xénero complementario é o de *colexial* a quen lle outorgan un feminino *colexiala*.

O numeral *un* co seu feminino *unha* non pode tampouco remitirse a ningún dos grupos existentes porque só flexiona o xénero. Para a súa implementación en XIADA decidímonos por introducilos como entradas separadas ás que se lles proporciona na subetiqueta a caracterización morfolóxica completa (*Numeral, cardinal, determinante/non determinante, masculino ou feminino dependendo do caso, singular*):

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	un	Ncams		un	N	si	volga_2004
xeral	principal	unha	Ncafs		un	N	si	volga_2004

Finalmente, polo que respecta á flexión de xénero e número, só nos quedan por comentar tres grupos: o G38, o G42 e o G43, todos eles inusitados lingüisticamente.

O G38, cuxo representante é AMBOS, emprégase en XIADA para caracterizar e xerar catro indefinidos (*ambos, entrambos, senllos e varios*) e mais dous pronomes (*nosoutros e vosoutros*), mentres que o G42 (representante TRESCENTOS), pola súa banda, coincide nas terminacións co G38 mais non cos valores desas terminacións:

Grupo: G38		
Representante: ambos		
terminación	subetiqueta	exemplo
os	masculino plural	<i>amb-os</i>
as	feminino plural	<i>amb-as</i>

Grupo: G42		
Representante: trescentos		
terminación	subetiqueta	exemplo
os	masculino singular/plural	<i>trescent-os</i>
as	feminino plural	<i>trescent-as</i>

Os lemas integrados no G38 só flexionan o xénero. Aínda que nun primeiro momento poidamos pensar que para *varios* existen as formas masculina e feminina singulares, *vario/varia*, debemos lembrar que como indefinido só pode usarse en plural, correspondendo a flexión de xénero e número ao adxectivo. Lembremos que a asociación de lema e categoría que realizamos en XIADA permítenos discriminar nas formas plurais as pertencentes ao Indefinido das correspondentes ao Adxectivo.

A diferenza entre o G38 e o G42 radica no valor da terminación *-os*; no primeiro caso só plural, mentres que no segundo presenta valor numérico ambiguo. A caracterización *singular e plural* da terminación masculina do G42 utilízamola única e exclusivamente para a expresión das centenas dos numerais cardinais no lexicón. Como xa vimos no apartado dedicado ao Numeral no Capítulo 1, consideramos que os cardinais non sempre indican número plural pois, por exemplo na expresión da

porcentaxe, o singular vén marcado pola concordancia co artigo. Confróntese no seguinte par de exemplos o masculino plural do primeiro co singular do segundo.

Os trescentos habitantes dunha illa deciden pór en marcha unha televisión local. [CORGA: C. Franco, *Palabras contadas*, Xerais, 2006]

A pesar diso teño que dicir que a relevancia da literatura para nenos de quince anos para aquí medrou un trescentos por cento. [CORGA: *VG1995-05-04/6*]

Esta ambigüidade no atributo número no lexicón non ten correlato na etiquetaxe de corpus, onde os valores singular e plural están sempre individualizados nos numerais. Ao G42 remítense 9 entradas, todas elas numerais que expresan as centenas a partir de *trescentos*. Entre esas entradas non figuran as correspondentes á segunda centena porque presentan cambios na raíz segundo o xénero, *douscentos* fronte a *duascentas*⁹⁰, mentres que si consta *quiñentos* e mais a non normativa *quinientos*.

O G43 fecha a relación de grupos de derivación que facilitan a introdución no lexicón de lemas que flexionan o xénero e o número. Sen dúbida, dá a sensación de que xa enumeramos todas as posibilidades, e en certo sentido así é. Este grupo non engade ningunha combinación diferente ás xa vistas nin ten representante, pois nel non finaliza ningunha análise, pero posibilita a caracterización e xeración dos superlativos sintéticos que se forman co sufixo *-ísim*. Na táboa que segue recollemos o seu contido e engadimos, para facilitar a lectura e comprensión, o contido do grupo ao que se deriva a terminación *-ísim*:

Grupo: G43						
<i>terminación</i>	<i>subetiqueta</i>	<i>seguinte grupo</i>	<i>normativa</i>	G1		<i>Exemplo</i>
				<i>terminación</i>	<i>subetiqueta</i>	
				o	ms (<i>masculino singular</i>)	<i>gravísimo</i>
ísim	superlativo	G1	si	a	fs (<i>feminino singular</i>)	<i>gravísimos</i>
				os	mp (<i>masculino plural</i>)	<i>gravísima</i>
				as	fp (<i>feminino plural</i>)	<i>gravísimas</i>

Nestes momentos no lexicón só se remiten 18 entradas ao grupo G43. A causa deste número tan reducido ten que ver coa decisión de non introducir sistematicamente no lexicón os superlativos, senón ilos integrando a medida que fosen aparecendo nos textos que se etiquetasen automática ou manualmente.

⁹⁰ A introdución no lexicón destas formas realizouse tratándoas como entradas separadas, unha para o masculino e outra para o feminino, aínda que, obviamente, cun lema común.

2.2.2.2.2. *A flexión de número*

Segundo Álvarez, Regueira e Monteagudo (1986: 64) “*a categoría de número marca a oposición entre o término non marcado, o singular, que designa unha especie ou un individuo da especie, e o término marcado, plural, que designa máis dun individuo da especie.*”

En galego o número márcase a través dun morfema flexivo que se engade sobre a forma de singular para obter a de plural de acordo cunhas regras determinadas. No sistema de XIADA, a semellanza do que viamos para a flexión do xénero e o número, non só se teñen en conta esas regras no establecemento dos distintos grupos, senón tamén calquera outra modificación gráfica á que puidera dar lugar a formación do plural.

Malia flexionar unicamente o número, seguen sendo grupos de derivación polo que non se altera a denominación. Así mesmo, a numeración dos grupos para a flexión de número é correlativa á de xénero e número, ocupando os grupos do G29 ao G37 e o G46. A que non resulta totalmente correlativa é a de xénero e número por térense establecido algúns dos grupos máis tarde. Cando os temos que describir, como agora, esta organización non é esteticamente a mellor; non obstante, para a estrutura do lexicón ou a aplicación do etiquetador non ten a menor importancia, razón pola que, no seu momento, preferimos non reorganizar os grupos, evitando deste xeito volver sobre o xa feito, posto que xa se estivera traballando na implementación do léxico.

Como estes grupos unicamente van acoller as terminacións e subetiquetas das entradas nas que se flexiona o número, non temos atoparnos cunha lista moi ampla. En realidade, 9 son as posibles combinacións mediante as que se produce a caracterización e xeración das formas variables en canto ao número no noso lexicón.

Tres desas combinacións coñecémolas todos; trátase das recollidas nos grupos G29, G30 e G32 correspondentes, respectivamente, á formación do plural mediante os alomorfos /s/, /es/ e /is/:

Grupo: G29			Grupo: G30			Grupo: G32		
Representante: alegría			Representante: mar			Representante: xenial		
terminación	subetiqueta	exemplo	terminación	subetiqueta	exemplo	terminación	subetiqueta	exemplo
	singular	<i>alegría</i>		singular	<i>mar</i>	l	singular	<i>xenia-l</i>
s	plural	<i>alegría-s</i>	es	plural	<i>mar-es</i>	is	plural	<i>xenia-is</i>

Ao G29 remitiáanse ata hai pouco as raíces que, sen sufrir ningunha modificación, engadían *-s* no plural, coma ALEGRÍA, representante do grupo. Mais aquí non só se caracterizaba a forma plural senón tamén a singular e nel tiñan cabida todos os substantivos ou adxectivos cuxo singular acababa tanto en vogal como en consoante *-n* ou calquera outra consoante distinta de *-l*, *-r*, *-z*, *-s* e *-x* que, para formar o plural, engadisen *-s*. A finais de 2010 decidimos acometer unha modificación neste grupo e extraer todas as raíces que rematasen en *-n* para trasladalas a outro grupo, o G46. O motivo que xustifica a creación dun novo grupo derivacional e alteracións noutro existente é proporcionar maior cobertura no recoñecemento e caracterización á hora de etiquetar os textos do CORGA. Nestes é bastante habitual para o plural das unidades rematadas en *-n* a presenza da variante dialectal correspondente ao bloque central: *canción / canciós, can / cas...* Esta é a razón de que se extraesen automaticamente todas as raíces que no G29 remataban en *-n* e se inclúsen no G46. Polo tanto, agora no G29 figuran raíces que, como o seu representante, engaden *-s* para formar o plural, exceptuando por suposto as voces rematadas en *-n* no singular.

A pesar de ser as clases Substantivo e Adxectivo as que máis elementos achegan ao G29, nel tamén están representadas categorías coma Pronome, Indefinido, Posesivo ou Artigo. Vexamos a táboa cos datos exactos:

24 709 entradas	24 706 lemas	Categoría	Exemplos
21 820	21 820	Substantivo	<i>abrazo, aceiro...</i>
2877	2877	Adxectivo	<i>abominable, abraiante...</i>
5	4	Posesivo	<i>meu, miña, mi, tu, su</i>
4	3	Pronome	<i>o, a, lle, vostede</i>
2	1	Artigo	<i>o, a</i>
1	1	Indefinido	<i>bastante</i>

Malia que, en teoría, no G29 o número de entradas debería traducirse estritamente no mesmo número de lemas, posto que nos substantivos e adxectivos que engaden *-s* para formar o plural a raíz e o lema coinciden, isto non é exactamente así debido a que para implementar algúns elementos das categorías Pronome, Artigo e Posesivo valémonos de raíces na súa forma feminina que se remiten ao lema na súa forma masculina. Dáse isto por exemplo nas formas femininas do artigo determinado e o pronome átono de terceira acusativo, *a*, que se remiten ao lema *o*, nun caso o Artigo e noutro o Pronome. Ou coa entrada posesiva *miña* que se remite ao lema *meu*.

O G46, pola súa banda, posúe as seguintes características:

Grupo: G46			
Representante: can			
<i>terminación</i>	<i>subetiqueta</i>	<i>normativa</i>	<i>exemplo</i>
n	singular	si	<i>ca-n, paixó-n</i>
ns	plural	si	<i>ca-ns, paixó-ns</i>
s	plural	non	<i>ca-s, paixó-s</i>

Tal como vemos, son tres as formas que caracteriza ou xera o grupo derivacional G46 por cada entrada incorporada: a singular (*canción*), a plural normativa (*cancións*) e a plural non normativa (*canciós*). Tras crear o grupo coas terminacións e subetiquetas correspondentes, extraéronse automaticamente do G29 todas as raíces que remataban en *-n* pertencentes ao lexicón *xeral* e incorporáronse ao G46 eliminándolles o *-n* final do singular, xa que este agora constitúe unha terminación. En total son 2957 raíces ou lemas os que engrosan o G46 desagregándose estes en 2939 substantivos (*acción, admisión, alacrán...*) e 18 adxectivos (*afín, ruín, palleirán...*).

O G30, cuxo representante é MAR, acolle todas as raíces que, sen alterarse de ningún xeito para formar o plural, engaden *-es*; é dicir, os monosílabos e polisílabos non agudos rematados en *-l* (*mel, ril, túnel, etc.*), as palabras compostas rematadas nun dos monosílabos anteriores (*ollomol, cartafol...*), as formas rematadas en *-r* (*mar, calor...*) e, finalmente, os monosílabos finalizados en *-s* (*gas, mes, gris, etc.*) ou polisílabos que non varían a raíz do singular ao plural (*adeus, adeuses* ou *país, países*).

Malia que puideramos pensar que o grupo G30, polas súas características, sería o máis frecuente despois do G29, os datos demostran que no noso lexicón ocupa a terceira posición, cos seguintes valores:

1385 entradas	1384 lemas	Categoría	Exemplos
828	828	Substantivo	<i>acendedor, automóbil</i>
551	551	Adxectivo	<i>angular, débil</i>
2	1	Interrogativo-exclamativo	<i>cal, cál</i>
1	1	Indefinido	<i>tal</i>
1	1	Pronome	<i>il</i>
1	1	Relativo	<i>cal</i>

A diferenza entre o número de entradas e o de lemas, 1, estriba na remisión do interrogativo-exclamativo *cál* ao lema sen acentuar, *cal*, pois nos interrogativos-exclamativos consideramos que as formas con til diacrítico, por estar o seu emprego, en teoría, unicamente limitado a evitar anfiboloxías, son unha variante, condicionada polo contexto, da forma sen acentuar. Cures, Mosquera e Negro (2005: 165-188) demostran que, na práctica, o tratamento ortográfico das denominadas oracións interrogativas indirectas no galego escrito moderno foi unha cuestión problemática e confusa para os usuarios da lingua escrita estándar nos distintos ámbitos.

Cando o alomorfo /es/ se lle engade a unha palabra polisílaba aguda rematada en -s prodúcese, á parte da simple oposición singular fronte a plural, unha diferenza gráfica entre a forma singular e a plural, provocada polas regras de acentuación, que empece a implementación destas palabras no G30. É o que ocorre, por exemplo, en *cortés* fronte a *cortesés*, mais tamén en *compás/compases* e *autobús/autobuses*. Computacionalmente, non seguen o modelo establecido para *mar*, pois nas formas plurais hai un cambio con respecto ás singulares: a perda do til gráfico. Isto provoca que, como mínimo, teñamos que crear un grupo novo para incluír nas terminacións o segmento que diverxe en ambos os pares. Con todo, non abonda con crear outro grupo, posto que, se consideramos raíz a parte común a cada par (*comp-*, *cort-* e *autob-*), as terminacións non nos permiten integralas no mesmo grupo pois non coinciden (-ás, -és, -ús), co que nos vemos obrigados a crear un grupo distinto para cada unha desas terminacións:

Grupo: G31			Grupo: G35			Grupo: G36		
Representante: cortés			Representante: compás			Representante: autobús		
terminación	subetiqueta	exemplo	terminación	subetiqueta	exemplo	terminación	subetiqueta	exemplo
és	singular	<i>cort-és</i>	ás	singular	<i>comp-ás</i>	ús	singular	<i>autob-ús</i>
eses	plural	<i>cort-eses</i>	ases	plural	<i>comp-ases</i>	uses	plural	<i>autob-uses</i>

Por suposto, as cifras demostran que son poucos os lemas que caracterizan no lexicón:

	34 lemas	Categoría	Exemplos
G31	16	Substantivo	<i>ciprés, revés</i>
	4	Adxectivo	<i>descortés, tarrantés</i>
G35	6	Substantivo	<i>augarrás, ananá</i>
	1	Adxectivo	<i>antigás</i>
G36	7	Substantivo	<i>aerobús, microbús</i>

Para rematar coas formas que constrúen o plural co alomorfo /es/ fáltanos aínda por comentar un grupo máis, o G34, cuxo representante é AUDAZ. A explicación da súa individualización xa a tratamos no establecemento do grupo G22, tipo RAPAZ, polo que non imos repetirnos. A diferenza entre eles estriba, loxicamente, en que no G22 flexiónase xénero e número, mentres que ao G34 só se remiten entradas que flexionan o número, ou sexa, caracterízanse e xéranse as formas masculinas ou femininas que cambian o *-c* final do singular por *-ces* plural:

Grupo: G34		
Representante: audaz		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplo</i>
z	singular	<i>auda-z</i>
ces	plural	<i>auda-ces</i>

Este modelo flexional de número é máis frecuente do que pensamos, pois a el remítense 206 lemas no lexicón de XIADA, distribuídos en 153 substantivos (*cariz*) e 53 adxectivos (*eficaz*).

Finalmente, o alomorfo /is/ emprégase nas palabras polisílabas agudas cuxo singular remata na consoante *-l*. Nestas, o *-l* final do singular é substituído no plural por *-is*, mais isto só sucede cando a vogal da sílaba oxítónica é “a”, “e” ou “o”: *animal*, *animais*; *papel*, *papeis*; *control*, *controis*. O G32 é o grupo encargado da caracterización e xeración das formas que cumpren o requisito anterior:

Grupo: G32		
Representante: xenial		
<i>terminación</i>	<i>subetiqueta</i>	<i>exemplos</i>
l	singular	<i>xenia-l</i> , <i>pape-l</i> , <i>contro-l</i>
is	plural	<i>xenia-is</i> , <i>pape-is</i> , <i>contro-is</i>

O modelo flexional numérico establecido no G32 é, como non podía ser doutra maneira, o que ocupa a segunda posición na listaxe de produtividade dos grupos que só flexionan o número no lexicón de XIADA, cunha cantidade total de lemas que case dobra o terceiro clasificado:

3017 lemas	Categoría	Exemplo
576	Substantivo	<i>animal, anzol, cardeal...</i>
2441	Adxectivo	<i>normal, banal, cerebral...</i>

O G33, cuxo representante de grupo é EDIL, pola súa banda, acolle os substantivos e adxectivos polisílabos agudos rematados en *-il*, aos que tamén se lles engade o alomorfo /is/, só que neste caso prodúcese a crase da vogal da base coa da do morfema de plural, orixinando a terminación *-ís*. Isto provoca que no grupo teñamos que trasladar xa no singular a vogal final da raíz á terminación:

Grupo: G33		
Representante: edil		
terminación	subetiqueta	exemplos
il	singular	<i>ed-il</i>
ís	plural	<i>ed-ís</i>

O grupo G33 acolle no noso lexicón 96 lemas, repartidos en 67 substantivos (*abril*) e 29 adxectivos (*caciquil*).

O derradeiro grupo desta longa enumeración en que se converteu a descrición formal dos modelos empregados computacionalmente para caracterizar e xerar as formas que flexionan xénero e/ou número en XIADA, o G37, só acubilla o substantivo REAL, singularizado no noso sistema porque presenta un plural non normativo, *reás*, bastante frecuente nos textos do CORGA⁹¹, que non pode remitirse a ningún dos grupos que xa vimos por non existir coincidencia nas terminacións:

Grupo: G37		
Representante: real		
terminación	subetiqueta	exemplos
al	singular	<i>re-al</i>
ás	plural	<i>re-ás</i>

2.2.2.2.3. O verbo. Grupos formais

É curioso constatar como unha única categoría, o verbo, require un amplo despregamento de modelos que acollan toda a súa complexidade formal. Por suposto, isto non se debe unicamente ás 70 formas que resultan da conxugación de cada verbo,

⁹¹ Na versión 1.4 documéntanse máis de 100 ocorrencias que abranguen temporalmente dende o ano 1975 ata o 2002. Dispoñible en <http://corpus.cirp.es/corga> [consultada o 23/07/2010].

senón á diferenza entre verbos regulares e verbos irregulares, clasificando nestes últimos non só os que presentan raíces distintas para o tema de presente e o de pretérito, senón todos aqueles que inclúen algunha variación ortográfica con respecto aos regulares.

Fronte aos 63 modelos distintos de conxugación verbal que recolle Díaz Regueiro (1992) nós conseguimos reducir esa cifra a 45: 3 modelos regulares e 42 irregulares. Agora ben, quizais sería máis apropiado falar de 90 modelos diferentes, posto que 45 están destinados a caracterizar as formas verbais do sistema galego sen clíticos engadidos ou segundas formas do artigo e os outros 45 foron creados para permitir o recoñecemento morfolóxico das distintas formas verbais, de calquera conxugación e calquera verbo, cando se constrúen con pronomes enclíticos ou se emprega a segunda forma do artigo.

Naturalmente, dada a estrutura do lexicón de XIADA, é posible tamén xerar a conxugación completa dos verbos que figuran no lexicón ou daqueles que escollamos, sen enclíticos ou cun, dous, tres ou catro enclíticos.

A necesidade de definir dous conxuntos conxugacionais diferentes, un susceptible de levar clíticos e outro sen eles, vén determinada pola ortografía da forma verbal en concorrencia con pronomes átonos enclíticos, pois estes gráfanse unidos directamente ao verbo e, a efectos de acentuación, o conxunto formado polo verbo e os pronomes forma un todo. Isto non quere dicir que nos grupos verbais do lexicón estean recollidos os pronomes átonos, senón simplemente que nos grupos verbais debemos dar conta das modificacións que se producen nas formas verbais por engadir enclíticos e/ou optar polo uso da segunda forma do artigo.

Así, por exemplo *cantarás*, prototipo da segunda persoa do singular do futuro de indicativo de calquera verbo regular da 1ª conxugación, se leva un pronome enclítico xa non se acentúa graficamente: *cantarasme*; se leva máis dun enclítico, sempre e cando non se trate dunha contracción monosilábica, volve recuperar o til: *cantarásllela* vs. *cantaraslla*. Mais non rematan aí os condicionantes porque, se ademais se quere construír coa segunda forma do artigo ou co pronome de terceira acusativo, amais de desaparecer o acento gráfico desaparece tamén o “s” final, morfema de número e persoa característico da segunda persoa singular: *cantara-la canción mañá* / *cantarala mañá*.

Neste apartado non imos reflectir o contido de todos os grupos verbais, senón que limitaremos a exemplificación de terminacións e subetiquetas á primeira conxugación regular. A reprodución íntegra dos 90 modelos ocuparía un espazo desmesurado e convertería a lectura deste capítulo en pesada, mentres que non achegaría información moi diferente da que se desprende do modelo que deseguido expoñeremos. Si daremos conta, non obstante, das características de todos os modelos verbais: que motiva a súa individualización, cantas e que raíces son precisas en cada modelo e cantos verbos do noso lexicón se agrupan en cada modelo. A maiores indicaremos tamén as correspondencias cos grupos verbais que levan enclíticos.

O sistema de denominación nos modelos regulares é sinxelo. *V* designa a clase de palabra Verbo e 1, 2 e 3 a clase verbal segundo a que se conxuga o verbo de que se trate. Exemplos:

V1 -> CANTAR

V2 -> COMER

V3 -> PARTIR

Estes tres grupos regulares, se se conxugan con enclíticos, pasan a denominarse

Vc1 -> CANTAR

Vc2 -> COMER

Vc3 -> PARTIR

onde “c” designa *con clíticos*.

Nos verbos irregulares o sistema de denominación é semellante, só que o número que aparece no nome do grupo non indica que o verbo se conxugue segundo a primeira, segunda ou terceira conxugación, senón o modelo no que nós o incluimos no lexicón para caracterizalo e conxugalo se así se desexa: Vi1, Vi2, Vi3, Vi4... Vi42. O “i” indica que se trata dun verbo irregular.

Os grupos que recollen o paradigma dos verbos irregulares con pronomes enclíticos reciben o mesmo nome ca os da conxugación sen clíticos, coa diferenza de que no nome do grupo, despois do “i” de irregular, aparece un “c” que se traduce en *con clíticos*: Vic1, Vic2, Vic3, Vic4... Vic42.

Todos os verbos regulares sen clíticos que se conxugan segundo a 1ª, 2ª ou 3ª conxugación presentan para a súa introdución no lexicón de XIADA unha única raíz: CANT-, COM-, PART- son as raíces pertinentes para introducir o paradigma completo de *cantar*, *comer* e *partir*. Esas raíces remítense, respectivamente, aos grupos V1, V2 e V3, onde se recollen as terminacións e se proporciona para cada unha delas, mediante a subetiqueta, a caracterización das 70 formas verbais que presenta a conxugación completa de calquera verbo regular galego: tempo, modo, persoa, xénero e número.

A excepción constitúena as terminacións “ad” e “id”, correspondentes á subetiqueta 0p0 (*tempo non aplica, participio, persoa non aplica*), pertencentes ás formas participiais, pois aproveitamos a existencia do grupo derivacional G1 (onde *o* = masculino singular; *a* = feminino singular; *os* = masculino plural; *as* = feminino plural), visto nunha epígrafe anterior, para remitir a el as desinencias “ad” e “id” e, deste xeito, completar a forma e etiqueta de todas as formas do participio: *cantado, cantada, cantados, cantadas / comido, comida, comidos, comidas / partido, partida, partidos, partidas*.

O participio flexiona xénero e número coma os substantivos ou adxectivos que seguen o modelo G1, polo que non ten razón de ser crear un grupo novo ou incluír as 4 formas participiais en cada un dos 45 modelos verbais que permiten a conxugación dos verbos sen clíticos. Así posibilitamos a caracterización e recoñecemento das catro formas do participio sen ter que incluílas unha a unha en cada modelo.

A continuación presentamos o contido do grupo V1 que inclúe as desinencias, subetiquetas, información sobre o grupo de derivación ao que se envía a terminación, se é pertinente, e normativa da forma. Con isto é posible analizar e xerar o paradigma verbal completo de calquera verbo regular da 1ª conxugación. Nel, como se observará, desinencia integra todo o que non é raíz, ou sexa, vogal temática, sufixo modotemporal e sufixo de número e persoa.

Malia que no V1 figuran as terminacións e as correspondentes subetiquetas sen ningunha orde predeterminada nin subclasificacións de ningún tipo, aquí, para a súa mellor comprensión, expoñerémolas indicando o tempo e modo que describen. A persoa despréndese facilmente da etiqueta (1, 2, 3) así como o xénero (0 = *non aplica*) e o número (s = singular; p = plural). Incluímos tamén a raíz (CANT-), non presente no V1,

para que non se perda de vista ao longo de todo o grupo e, ademais, a modo de exemplo, as formas verbais que se definen, aínda que estas non están no grupo V1 nin no lexicón, senón que hai que xeralas a partir da información que se desprende da raíz e grupo ao que se deriva.

Tras as aclaracións anteriores, necesarias para entender e localizar onde corresponde cada elemento das táboas, vexamos, pois, cal é o contido do grupo V1:

		Presente de indicativo				
		<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-		o	pi10s		si	<i>canto</i>
		as	pi20s		si	<i>cantas</i>
		a	pi30s		si	<i>canta</i>
		amos	pi10p		si	<i>cantamos</i>
		ades	pi20p		si	<i>cantades</i>
		an	pi30p		si	<i>cantan</i>

No presente de subxuntivo, dada a coincidencia formal entre a primeira e a terceira persoa singular, empregamos no atributo persoa o valor ambiguo que as aglutina. Á hora de etiquetar corpus ou extraer o léxico de XIADA, un *script* converte esa etiqueta ambigua en tres distintas, unha correspondente á primeira persoa (Vps10s), outra pertencente á terceira (Vps30s) e, finalmente, outra etiqueta na que permanece a ambigüidade sen resolver (Vpsa0s). Isto é aplicable a todos os casos nos que se emprega o valor aglutinador *a* para o atributo persoa:

		Presente de subxuntivo				
		<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-		e	psa0s		si	<i>cante</i>
		es	ps20s		si	<i>cantes</i>
		emos	ps10p		si	<i>cantemos</i>
		edes	ps20p		si	<i>cantedes</i>
		en	ps30p		si	<i>canten</i>

No copretérito, pospretérito e antepretérito de indicativo volve aparecer o valor *a*, aglutinador da primeira e terceira persoa do singular, para traducir o sincretismo formal existente nesas persoas:

		Copretérito de indicativo				
		<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-		aba	iia0s		si	<i>cantaba</i>
		abas	ii20s		si	<i>cantabas</i>
		abamos	ii10p		si	<i>cantabamos</i>
		abades	ii20p		si	<i>cantabades</i>
		aban	ii30p		si	<i>cantaban</i>

Pospretérito de indicativo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	aría	cia0s		si	<i>cantaría</i>
	arías	ci20s		si	<i>cantarías</i>
	ariamós	ci10p		si	<i>cantariamos</i>
	ariades	ci20p		si	<i>cantariades</i>
	arían	ci30p		si	<i>cantarían</i>

Antepretérito de indicativo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ara	lia0s		si	<i>cantara</i>
	aras	li20s		si	<i>cantaras</i>
	aramós	li10p		si	<i>cantaramos</i>
	arades	li20p		si	<i>cantarades</i>
	aran	li30p		si	<i>cantaran</i>

Pretérito de indicativo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ei	ei10s		si	<i>cantei</i>
	aches	ei20s		si	<i>cantaches</i>
	ou	ei30s		si	<i>cantou</i>
	amos	ei10p		si	<i>cantamos</i>
	astes	ei20p		si	<i>cantastes</i>
	aron	ei30p		si	<i>cantaron</i>

No pretérito de subxuntivo danse os únicos casos do paradigma verbal nos que unha desinencia está caracterizada como non normativa. Trátase, no grupo V1, das terminacións graves “asemos” e “asedes” do pretérito de subxuntivo, as cales se rexeitan na edición do 2003 das *Normas* (RAG/ILG, 2003) en favor das terminacións esdrúxulas “ásemos” e “ásedes”:

Pretérito de subxuntivo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ase	esa0s		si	<i>cantase</i>
	ara	esa0s		si	<i>cantara</i>
	ases	es20s		si	<i>cantases</i>
	aras	es20s		si	<i>cantaras</i>
	ásemos	es10p		si	<i>cantásemos</i>
	asemos	es10p		non	<i>cantasemos</i>
	aramós	es10p		si	<i>cantaramos</i>
	ásedes	es20p		si	<i>cantásedes</i>
	asedes	es20p		non	<i>cantasedes</i>
	arades	es20p		si	<i>cantarades</i>
	asen	es30p		si	<i>cantasen</i>
	aran	es30p		si	<i>cantaran</i>

As formas en *-ra* subxuntivas puxéronnos nun dilema, pois na maior parte dos estudos divulgativos relativos ao sistema verbal galego omítese na presentación e nos esquemas da conxugación a forma en *-ra* para o pretérito de subxuntivo, co que, en

consecuencia, non se lle recoñece o valor subxuntivo (Díaz Regueiro, 1992; Nicolás Rodríguez, 1993; González et al., 2002; o conxugador verbal do *Diccionario de Galego* de Ir Indo⁹², etc.); noutros casos deféndese explicitamente que se trata dunha incorrección (Freixeiro, 2000: 358; Hermida, 2004: 110-112).

O único texto prescriptivo vixente do galego, as *Normas ortográficas e morfolóxicas do idioma galego* (RAG/ILG, 2003: 111), recomenda “*manter ben discriminados os usos dos dous tempos (-ra coma antepretérito de indicativo e -se coma pretérito de subxuntivo)*”.

Fronte á omisión ou consideración xeral de que a forma en *-ra* non pode funcionar como subxuntiva, Álvarez e Xove⁹³ (2002: 294) defenden que o “*Pto. realízase con dúas variantes de expresión, |ra| e |se|, que ocorren tanto no Subx. non irreal coma no Subx. irreal. Son variantes máis ou menos libres, que en principio se poden empregar indistintamente, mais hai diferencias dialectais e de significado modal entre elas: a) nalgunhas zonas, sobre todo nas máis occidentais, a forma máis empregada e mesmo única é |ra|; pola contra, na zona oriental emprégase máis |se| ca |ra|; b) se o Pto. é “non-irreal” aumenta proporcionalmente o uso de |se|, e se é irreal medra a frecuencia de |ra|*”.

As investigacións dun especialista no sistema verbal latino, castelán e galego como é Alexandre Veiga avalan a vixencia e corrección en galego dos usos subxuntivos das formas en *-ra*:

No referente ó galego, a situación de *cantara* é parcialmente semellante á do castelán antigo e parcialmente semellante á do moderno: como no primeiro, *cantara* funciona como *antepretérito* (“pluscuamperfecto”) de IND Ø; como no segundo, funciona como alomorfo de *cantase* en SUBX Ø e SUBX 2 e conserva usos residuais –tamén fundamentalmente cos verbos *poder, deber, querer*– (cfr. supra §§ 1, 2.1) do seu emprego como IND 2. O resultado é que en galego moderno *cantara* aparece nas catro funcións modais resultantes da combinación das oposicións indicativo / subxuntivo e ± irreal:

IND Ø: conteille que *foras* ti quen mo *dixera*

⁹² Disponible en <http://www.digalego.com>. Cando se consulta unha forma verbal o diccionario ofrece a posibilidade de comprobar como se conxuga todo o paradigma, por exemplo *andar*, e nel non aparece tampouco a forma en *-ra* con valor subxuntivo. [Consultado o 21/06/2010].

⁹³ Anos antes, o propio Xove (1982: 319) afirmaba que “*a forma cantara é, diacrónica e sincronicamente, un antepretérito de indicativo, pro na narrativa de Cunqueiro aparece tamén cos mesmos significados e distribución có pretérito de subxuntivo*”.

IND 2: *quixera* [~ *querería* ~ *quería*] pedirlle un favor

SUBX Ø: mandeille que o *fixera* (~*se*)

SUBX 2: oxalá *chovera* (~*se*) [Veiga, 1991: 525-526]

É máis, Veiga refuta a presenza das formas en *-ra* no pretérito de subjuntivo como castelanismo, e conclúe:

No es verosímil suponer que el surgimiento en gallego de los usos subjuntivos de la forma verbal *cantara* se haya debido a ninguna presión de la lengua castellana sobre la gallega. De nuevo prudentemente, no descartaremos que su generalización en gallego hasta los extremos de haber hecho desaparecer a la forma etimológicamente subjuntiva *cantase* en alguna variedad geográfica concreta hubiera podido verse facilitada por el paralelo triunfo de *-ra* sobre *-se* en español, pero en modo alguno puede situarse en el verbo castellano la causa del proceso. [Veiga, 2005: 23-50]

Concordamos, pois, coas análises propostas por Álvarez e Xove (2002) e Veiga (1991, 1993, 2005) para o verbo galego, e por iso nos nosos modelos conxugacionais mantemos para o pretérito de subjuntivo as formas en *-ra* como normativas. Paradoxalmente, isto provoca que na conxugación do mesmo tempo verbal, o pretérito de subjuntivo, coexistan xa non só dúas expresións para un mesmo contido modotemporal (*/ra/* e */se/*), senón tamén dúas estruturas acentuais diverxentes para a primeira e segunda persoas do plural: unha paroxítona, *cantaramos*, *cantasemos* fronte a outra proparoxítona, *cantásemos*.

Por suposto, recoñecemos que o cruzamento entre a dupla expresión modotemporal, a dobre acentuación e a caracterización da normatividade achega uns resultados, como mínimo, rechamantes, pois

- 1) as formas en *-ra* graves son normativas: *cantaramos*, *cantarades*.
- 2) as formas en *-se* graves son non normativas: **cantasemos*, **cantasedes*.
- 3) as formas en *-se* esdrúxulas son normativas: *cantásemos*, *cantásedes*.

4) as formas en *-ra* esdrúxulas non están incorporadas nos modelos de XIADA, mais se o estivesen serían non normativas⁹⁴: **cantáramos*, **cantárades*

⁹⁴ A acentuación proparoxítona na primeira e segunda do plural do copretérito, pospretérito e antepretérito de indicativo, e mais no pretérito de subjuntivo considérase un trazo dialectal ata a edición das *Normas* do 2003 (RAG/ILG, 2003), polo que non se rexistra nos modelos conxugacionais dos distintos manuais

Por que se produce esta situación? A súa orixe está na pretensión de conxugar as normas oficiais do galego coa tese defendida entre outros por Veiga (1991, 1993, 2005), e que nós compartimos, xunto coa necesidade de construír e adestrar ferramentas que se empregarán para a análise morfolóxica dun corpus de galego contemporáneo. A secuenciación do problema podería ser a seguinte:

1º) Oficialmente non se recoñecen as formas en *-ra* para o pretérito de subxuntivo, non obstante estas si se empregan na fala e nos documentos do corpus, motivo polo que as introducimos nos paradigmas verbais do lexicón.

2º) Como as caracterizamos normativamente? Segundo a normativa oficial non son correctas, mais hai outros estudos que demostran que as formas en *-ra* do pretérito de subxuntivo non son produto da interferencia do castelán, senón resultado da evolución do sistema verbal, así que decidimos non seguir as *Normas* (RAG/ILG, 2003) e consideralas normativas. A estas alturas todas as primeiras e segundas persoas do plural do paradigma verbal son graves; a única incongruencia que presenta o noso lexicón coa normativa oficial é a de concederlles ás formas en *-ra* subxuntivas o estatus de normativas.

3º) A nova edición das *Normas* (RAG/ILG, 2003: 107-108) sorpréndenos cun cambio inesperado, pois malia recoñecer que a acentuación paroxítóna na 1ª e 2ª persoa do plural é “*un trazo conservador que se debe preservar na lingua normativa, dado que non se trata de ningunha reliquia, senón dunha pronuncia moi común aínda, que ocupa máis da metade sur de Galicia, e mesmo non é allea á parte norte*”; malia que “*a regresión das formas cantabamos, cantabades etc. débese en boa parte a que se toma como modelo de corrección o castelán e, en consecuencia, estas reprímense como vulgares*”, malia todo isto, os académicos indican que “*no pretérito de subxuntivo prefírense as formas proparoxítonas cantásemos e cantásedes*” e xustificano simplemente con que son “*maioritarias xa no uso moderno*”⁹⁵.

do verbo galego nin nos do noso lexicón. A modificación introducida pola última edición das *Normas* (RAG/ILG, 2003) altera substancialmente este panorama e provoca a situación que estamos a describir.

⁹⁵ Non podemos deixar de chamar a atención sobre tres paradoxos observados neste parágrafo das *Normas* (RAG/ILG, 2003: 107-108): 1º, consideración da acentuación esdrúxula como unha interferencia do castelán, mais aínda así prefírese para o pretérito de subxuntivo; 2º, xustificación polo seu uso maioritario na actualidade, cando ese mesmo uso avala tamén as formas en *-ra*; 3º, se a acentuación grave é unha reliquia no pretérito de subxuntivo, que papel desempeña o futuro de subxuntivo no paradigma verbal? Finalmente unha pregunta: é aceptable comparar a asimetría acentual do sistema verbal italiano

4º) Consecuentemente, introducimos nos paradigmas verbais do lexicón as formas en *-se* esdrúxulas e modificamos a indicación da normatividade nas correspondentes graves. Mais, que facer coas formas en *-ra*? En primeiro lugar, a Academia non as inclúe cando trata a acentuación das primeiras e segundas persoas do plural do pretérito de subxuntivo; en segundo lugar, no apartado pertencente a “Andase, andara” (RAG/ILG, 2003: 111), onde recoñece que as formas en *-ra* nalgúns casos son as únicas posibles, non se pronuncia sobre a acentuación da P4 e a P5 nin recolle exemplos desas persoas; en terceiro lugar, sería unha solución pouco acertada pola nosa banda pretender unha acentuación proparoxítona para elas cando a mesma expresión, aínda que cun contido modotemporal diferente (antepretérito de indicativo), é grave. Cal é a solución entón? Non tocarlles, deixalas como formas graves normativas, pois ou as consideramos non normativas, co que non estamos en absoluto de acordo, ou non está nas nosas mans corrixir esta anomalía.

Futuro de indicativo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	arei	fi10s		si	<i>cantarei</i>
	arás	fi20s		si	<i>cantarás</i>
	ará	fi30s		si	<i>cantará</i>
	aremos	fi10p		si	<i>cantaremos</i>
	aredes	fi20p		si	<i>cantaredes</i>
	arán	fi30p		si	<i>cantarán</i>

Futuro de subxuntivo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ar	fsa0s		si	<i>cantar</i>
	ares	fs20s		si	<i>cantares</i>
	amos	fs10p		si	<i>cantamos</i>
	ardes	fs20p		si	<i>cantardes</i>
	aren	fs30p		si	<i>cantaren</i>

Imperativo					
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	a	0m20s		si	<i>canta</i>
	ade	0m20p		si	<i>cantade</i>

No infinitivo e xerundio a primeira terminación corresponde á forma non persoal e as demais ás persoais, diferenciándose unhas das outras na persoa e número: nas formas non persoais non aplica ningún destes atributos.

co do galego para apoiar o desprazamento acentual á vogal temática da primeira e segunda persoa do plural do pretérito de subxuntivo en galego? Son comparables ambos os sistemas?

	Infinitivo				
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ar	0f000		si	<i>cantar</i>
	ar	0fa0s		si	<i>cantar</i>
	ares	0f20s		si	<i>cantares</i>
	armos	0f10p		si	<i>cantarmos</i>
	ardes	0f20p		si	<i>cantardes</i>
	aren	0f30p		si	<i>cantaren</i>

A pesar de que habitualmente as formas do xerundio conxugado non aparecen recollidas nos paradigmas verbais (Díaz Regueiro, 1992; RAG/ILG, 1995 e 2003; Nicolás Rodríguez, 1993, González et al., 2002; etc.), a súa presenza nos textos obríganos a darlles cabida nos distintos grupos verbais co fin de poder recoñecelas e caracterizalas axeitadamente. No que respecta á normatividade das formas correspondentes á primeira e segunda persoas do plural do xerundio conxugado, dado que no único texto preceptivo do galego actual, as *Normas* (RAG/ILG, 2003), non se fai referencia ningunha a este respecto, consultamos a cuestión co profesor Manuel González González⁹⁶, quen nos indicou que as podíamos tratar como normativas, polo que así o fixemos.

	Xerundio				
	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Seguinte grupo</i>	<i>Normativa</i>	<i>Exemplo</i>
CANT-	ando	0x000		si	<i>cantando</i>
	ándomos	0x10p		si	<i>cantándomos</i>
	ándodes	0x20p		si	<i>cantándodes</i>

A terminación do participio é a única, como explicamos máis arriba, que *salta* a outro grupo de derivación, o G1, polo que a súa caracterización morfolóxica non remata no V1 senón no G1. Incluímos na táboa exemplificativa o contido do grupo derivacional G1 co fin de que se perciba con claridade a *ruta* completa que seguen as formas do participio no sistema de XIADA mais, insistimos, as terminacións e subetiquetas do G1 non están incluídas no grupo verbal V1:

⁹⁶ Naquel momento, ano 2009, secretario da Real Academia Galega e director do Seminario de Lexicografía da RAG e do Servizo de Terminoloxía Galega (TERMIGAL), así como coautor do VOLGa, polo que o consideramos persoa totalmente cualificada para determinar a normatividade das formas do xerundio conxugado.

CANT-	Participio					
	Desinencia	Subetiqueta	Seguinte grupo	Normativa	G1	
ad	0p0	G1	si	terminación	subetiqueta	
				o	ms (masculino singular)	cantado
				a	fs (feminino singular)	cantada
				os	mp (masculino plural)	cantados
				as	fp (feminino plural)	cantadas

2.2.2.2.4. O verbo: a conjugación regular con enclíticos. Grupos formais

Fronte a unha única raíz nos verbos regulares sen clíticos que reside no sublexicón *principal*, cando estas formas conxugadas se constrúen con pronomes enclíticos ou coa segunda forma do artigo, a pesar de ser verbos totalmente regulares, debemos introducir no sublexicón *clíticos* dúas raíces distintas que se remiten a dous grupos diferentes.

As raíces átonas, CANT-, COM-, PART-, envíanse respectivamente aos grupos Vc1a, Vc2a e Vc3a, onde se subministra a caracterización de todas as formas verbais que levan o acento, gráfico ou non, na vogal tónica ou na desinencia modotemporal.

As raíces tónicas, con acento na vogal radical, CÁNT-, CÓM-, PÁRT-, remítense aos grupos Vc1at, Vc2at e Vc3at, respectivamente, nos que se recollen as terminacións e subetiquetas correspondentes ás primeiras, segundas e terceiras persoas singulares do presente de indicativo e subxuntivo e mais a segunda singular do imperativo. O *t* no nome do grupo indica que a raíz se acentúa graficamente. Na medida do posible procuramos reflectir con esta denominación a correspondencia que se establece na conxugación con clíticos entre raíces átonas e raíces tónicas:

Conxugación regular con enclíticos			
Raíz átona		Raíz tónica	
Grupo	Raíz	Grupo	Raíz
Vc1a	CANT-	Vc1at	CÁNT-
Vc2a	COM-	Vc2at	CÓM-
Vc3a	PART-	Vc3at	PÁRT-

A modo de exemplo imos ver o contido dos grupos Vc1a e Vc1at. Non debe sorprendernos no interior dos dous grupos a presenza de dúas terminacións coa mesma subetiqueta para as formas verbais rematadas en *-s* ou *-r*, pois dita terminación é necesaria para caracterizar e recoñecer esas formas verbais cando levan segunda forma do artigo ou o pronome átono de terceira acusativo *lo, la, los, las*.

Así, coa desinencia *-ámos*, por exemplo, reflectimos a análise de *cantamos*, primeira persoa do plural do presente ou pretérito de indicativo, cando leva un ou máis pronomes enclíticos, sempre que o pronome non sexa o de terceira acusativo ou a segunda forma do artigo: *cantámoscha, cantámosllela*, etc.

A terminación *-ámo* facilita o recoñecemento e xeración tamén da primeira persoa do plural do presente e pretérito de indicativo cando esta se constrúe con segunda forma do artigo ou co alomorfo *lo, la, los, las* do pronome de terceira acusativo ou o pronome de primeira plural acusativo/dativo *nos*: *Lavámo-la cara / Lavámola / Lavámonos*.

Igual que sucedía na conxugación sen clíticos, nos grupos Vc1a e Vc1at, segundo os que se caracterizan todas as formas verbais da 1ª conxugación regular cando concorren enclíticos, as terminacións e as correspondentes subetiquetas figuran sen ningunha orde predeterminada nin subclasificacións de ningún tipo; aquí, non obstante, para a súa mellor comprensión expoñerémolas indicando o tempo e modo que describen. A persoa despréndese facilmente da etiqueta (1, 2, 3) así como o xénero (0 = *non aplica*) e o número (s = singular; p = plural). Aínda que esta información non consta no lexicón, a modo de exemplo, incluimos ademais unha das posibles ocorrencias da forma verbal que se define acompañada dalgún ou varios clíticos, segundo corresponda.

Eliminamos tamén o campo *seguinte grupo* das táboas exemplificativas, pois na conxugación con clíticos en ningún momento é pertinente para caracterizar ningunha forma verbal, xa que o participio nunca se combina con enclíticos. Finalmente, pensamos que era preferible incluír na mesma táboa, a efectos de representación, o contido dos grupos Vc1a e Vc1at, converténdose deste xeito o tempo verbal en eixe central de cada táboa.

Comezaremos co presente de indicativo:

Presente de indicativo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	ámos	pi10p	si	<i>cantámoscha</i>
		ámo	pi10p	si	<i>cantámola</i>
		ádes	pi20p	si	<i>cantádesme</i>
		áde	pi20p	si	<i>cantádela</i>
[CÁNT]	Vc1at	o	pi10s	si	<i>cántoa</i>
		as	pi20s	si	<i>cántaslla</i>
		a	pi20s	si	<i>cántala</i>
		a	pi30s	si	<i>cántaa</i>
		an	pi30p	si	<i>cántannos</i>

Na conxugación con clíticos, a pesar de manterse o sincretismo formal entre a primeira e terceira persoa do singular no presente e futuro de subxuntivo, copretérito, pospretérito e antepretérito de indicativo e infinitivo conxugado, non mantemos a ambigüidade na persoa entre a 1ª e a 3ª singulares no lexicón, como si facemos na conxugación sen clíticos, porque a súa combinación con pronomes enclíticos esixe que estean separadas. Por moito que a forma *cantaría*, por exemplo, poida ser caracterizada como 1ª ou 3ª persoa do singular do pospretérito, se leva enclítico o pronome *se*, xa só é posible a súa análise como 3ª persoa. Por outra banda, as regras externas ao lexicón que regulan a combinación de certos clíticos con formas verbais precisan telas diferenciadas.

Presente de subxuntivo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	émos	ps10p	si	<i>cantémoscha</i>
		émo	ps10p	si	<i>cantémola</i>
		édes	ps20p	si	<i>cantédesme</i>
		éde	ps20p	si	<i>cantédela</i>
[CÁNT]	Vc1at	e	ps10s	si	<i>cánteche</i>
		es	ps20s	si	<i>cánteslla</i>
		e	ps20s	si	<i>cántela</i>
		e	ps30s	si	<i>cántese</i>
		en	ps30p	si	<i>cántennos</i>

Excepto no imperativo, onde son precisos os dous grupos, o Vc1a e o Vc1at, no resto de formas verbais que nos quedan por exemplificar –copretérito, pospretérito, antepretérito e futuro de indicativo, así como pretérito e futuro de subxuntivo ou infinitivo e xerundio– cando concorren con enclíticos, a raíz é sempre átona (CANT-) e remítese ao grupo Vc1a:

Copretérito de indicativo					
<i>[Raíz]</i>	<i>Grupo</i>	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Normativa</i>	<i>Exemplo</i>
[CANT]	Vc1a	ába	ii10s	si	<i>cantábache</i>
		ábas	ii20s	si	<i>cantábaslla</i>
		ába	ii20s	si	<i>cantábala</i>
		ába	ii30s	si	<i>cantábase</i>
		abámos	ii10p	si	<i>cantabámoscha</i>
		abámo	ii10p	si	<i>cantabámola</i>
		abádes	ii20p	si	<i>cantabádeslle</i>
		abáde	ii20p	si	<i>cantabádela</i>
		ában	ii30p	si	<i>cantábana</i>

Pospretérito de indicativo					
<i>[Raíz]</i>	<i>Grupo</i>	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Normativa</i>	<i>Exemplo</i>
[CANT]	Vc1a	aría	ci10s	si	<i>cantaríache</i>
		arías	ci20s	si	<i>cantaríaslla</i>
		aría	ci20s	si	<i>cantaríaala</i>
		aría	ci30s	si	<i>cantaríaase</i>
		ariámos	ci10p	si	<i>cantariámoscha</i>
		ariámo	ci10p	si	<i>cantariámola</i>
		ariádes	ci20p	si	<i>cantariádeslle</i>
		ariáde	ci20p	si	<i>cantariádela</i>
				arían	ci30p

Antepretérito de indicativo					
<i>[Raíz]</i>	<i>Grupo</i>	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Normativa</i>	<i>Exemplo</i>
[CANT]	Vc1a	ára	li10s	si	<i>cantáracha</i>
		áras	li20s	si	<i>cantáraslla</i>
		ára	li20s	si	<i>cantárala</i>
		ára	li30s	si	<i>cantárase</i>
		arámos	li10p	si	<i>cantarámoscha</i>
		arámo	li10p	si	<i>cantarámola</i>
		arádes	li20p	si	<i>cantarádeslle</i>
		aráde	li20p	si	<i>cantarádela</i>
				áran	li30p

No pretérito de indicativo proporciónanse dúas terminacións para a primeira e terceira persoa do singular, xa que dependendo do número de enclíticos silábicos que se lle engadan variará a acentuación do conxunto: *Canteina* ~ *cantouna moi mal* / *Cantéivola* ~ *cantóuvola moitas veces cando erades pequenos*.

Pretérito de indicativo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	éi	ei10s	si	<i>cantéivola</i>
		ei	ei10s	si	<i>canteina</i>
		áches	ei20s	si	<i>cantácheslles</i>
		áche	ei20s	si	<i>cantáchelas</i>
		óu	ei30s	si	<i>cantóullela</i>
		ou	ei30s	si	<i>cantouna</i>
		ámos	ei10p	si	<i>cantámoslle</i>
		ámo	ei10p	si	<i>cantámola</i>
		ástes	ei20p	si	<i>cantástesma</i>
		áste	ei20p	si	<i>cantástela</i>
		áron	ei30p	si	<i>cantáronlla</i>

O pretérito de subxuntivo é o tempo que máis terminacións acumula no paradigma verbal debido á presenza das formas en *-ra* e en *-se*:

Pretérito de subxuntivo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	áse	es10s	si	<i>cantásella</i>
		ára	es10s	si	<i>cantáraa</i>
		áses	es20s	si	<i>cantáseslle</i>
		áse	es20s	si	<i>cantásela</i>
		áras	es20s	si	<i>cantárasllas</i>
		ára	es20s	si	<i>cantáralas</i>
		áse	es30s	si	<i>cantáseas</i>
		ára	es30s	si	<i>cantáraas</i>
		ásemos	es10p	si	<i>cantásemoslla</i>
		ásemo	es10p	si	<i>cantásemola</i>
		asémos	es10p	non	<i>cantasémosllela</i>
		asémo	es10p	non	<i>cantasémola</i>
		arámos	es10p	si	<i>cantarámosvos</i>
		arámo	es10p	si	<i>cantarámola</i>
		ásedes	es20p	si	<i>cantásedeslla</i>
		ásede	es20p	si	<i>cantásedela</i>
		asédes	es20p	non	<i>cantasédeslla</i>
		aséde	es20p	non	<i>cantasédela</i>
		arádes	es20p	si	<i>cantarádeslla</i>
		aráde	es20p	si	<i>cantarádela</i>
		ásen	es30p	si	<i>cantásenlla</i>
		áran	es30p	si	<i>cantáranlla</i>

Malia que se adoita definir o modo subxuntivo como o modo da subordinación (Álvarez, Regueira e Monteagudo, 1986: 374), sendo nesta de regra a próclise do pronome, o subxuntivo, presente ou pretérito, pode aparecer en construcións non subordinadas como son as alternancias concesivas, polo que é posible, aínda que non moi frecuente, atopar pronomes enclíticos tanto no presente como no pretérito de subxuntivo:

Cantáraas ou non as cantara, iannos gañar igual.

Cantarádeslle *a canción ou non lla cantarades, ía seguir chorando.*

Futuro de indicativo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	aréi	fi10s	si	<i>cantaréivola</i>
		arei	fi10s	si	<i>cantareina</i>
		arás	fi20s	si	<i>cantarásllelas</i>
		aras	fi20s	si	<i>cantarasme</i>
		ara	fi20s	si	<i>cantarala</i>
		ará	fi30s	si	<i>cantarállela</i>
		ara	fi30s	si	<i>cantaraa</i>
		arémos	fi10p	si	<i>cantarémoslle</i>
		arémo	fi10p	si	<i>cantarémonos</i>
		arédes	fi20p	si	<i>cantarédeslle</i>
		aréde	fi20p	si	<i>cantarédelo</i>
		arán	fi30p	si	<i>cantaránvolas</i>
		aran	fi30p	si	<i>cantarana</i>

Recollemos deseguida o futuro de subxuntivo como un tempo verbal máis ao que, en teoría, é posible engadirlle algún clítico, aínda que non documentamos ningún exemplo e recoñecemos que para nós é totalmente artificial⁹⁷. As desinencias e exemplos son os mesmos ca os do infinitivo conxugado pois estes dous tempos presentan sincretismo formal na conxugación regular:

Futuro de subxuntivo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	ar	fs10s	si	<i>cantarlle</i>
		ár	fs10s	si	<i>cantárvola</i>
		a	fs10s	si	<i>cantala</i>
		áres	fs20s	si	<i>cantáresllas</i>
		áre	fs20s	si	<i>cantárela</i>
		ar	fs30s	si	<i>cantarlle</i>
		ár	fs30s	si	<i>cantárvola</i>
		a	fs30s	si	<i>cantala</i>
		ármos	fs10p	si	<i>cantármoslle</i>
		ármola	fs10p	si	<i>cantármola</i>
		árdes	fs20p	si	<i>cantárdeslle</i>
		árde	fs20p	si	<i>cantárdela</i>
		áren	fs30p	si	<i>cantárenllela</i>

Imperativo					
[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa	Exemplo
[CANT]	Vc1a	áde	0m20p	si	<i>cantádenola</i>
[CÁNT]	Vc1at	a	0m20s	si	<i>cántallela</i>

⁹⁷ Á parte de empregarse só en estilos cultistas e moi marcados (Álvarez e Xove; 2002), o seu uso está condicionado, sobre todo, ás oracións condicionais, modais e temporais, co que é de rigor a próclise e non a énclice.

Infinitivo					
<i>[Raíz]</i>	<i>Grupo</i>	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Normativa</i>	<i>Exemplo</i>
[CANT]	Vc1a	ar	0f000	si	<i>cantarlle</i>
		ár	0f000	si	<i>cantárvola</i>
		a	0f000	si	<i>cantala</i>
		ar	0f10s	si	<i>cantarlle</i>
		ár	0f10s	si	<i>cantárvola</i>
		a	0f10s	si	<i>cantala</i>
		áres	0f20s	si	<i>cantáresllas</i>
		áre	0f20s	si	<i>cantárela</i>
		ar	0f30s	si	<i>cantarlle</i>
		ár	0f30s	si	<i>cantárvola</i>
		a	0f30s	si	<i>cantala</i>
		ármos	0f10p	si	<i>cantármoslle</i>
		árm	0f10p	si	<i>cantármola</i>
		árdes	0f20p	si	<i>cantárdeslle</i>
		árde	0f20p	si	<i>cantárdela</i>
áren	0f30p	si	<i>cantárenllela</i>		

Dada a infrecuente aparición do xerundio conxugado nos textos e a inexistencia ata o momento de formas persoais xerundiais con algún clítico no CORGA, non cremos oportuno incluír sistematicamente no lexicón de XIADA as caracterizacións correspondentes ao xerundio persoal con pronomes enclíticos, quedándonos finalmente só coa forma non persoal:

Xerundio					
<i>[Raíz]</i>	<i>Grupo</i>	<i>Desinencia</i>	<i>Subetiqueta</i>	<i>Normativa</i>	<i>Exemplo</i>
[CANT]	Vc1a	ándo	0x000	si	<i>cantándovolas</i>

O grupo V1 que acolle a conxugación dos verbos da denominada Clase I ou primeira conxugación, modelo CANTAR, é con moito o máis produtivo pois a el remítense 4422 lemas.

A segunda conxugación regular ou Clase II está representada no lexicón de XIADA polo verbo COMER, modelo V2 segundo o que se conxugan 190 verbos.

Da Clase III ou terceira conxugación regular, cuxo representante é PARTIR, figuran en XIADA 223 lemas.

A modo de resumo recolleemos nunha táboa os datos relativos ás conxugacións regulares: raíces, grupos e número de lemas por cada modelo.

Conxugación regular							
Conxugación regular sen enclíticos			Conxugación regular con enclíticos				
			Raíz átona		Raíz tónica		
Clase	Grupo	Raíz	Grupo	Raíz	Grupo	Raíz	Lemas
I (<i>cantar</i>)	V1	CANT-	Vc1a	CANT-	Vc1at	CÁNT-	4422
II (<i>comer</i>)	V2	COM-	Vc2a	COM-	Vc2at	CÓM-	190
III (<i>partir</i>)	V3	PART-	Vc3a	PART-	Vc3at	PÁRT-	223

O número de verbos regulares que figuran no lexicón, 4835 en total, constitúe algo máis do dobre da totalidade de verbos irregulares que tamén rexistramos en XIADA, 2068, co que en conxunto caracterizamos no noso sistema o paradigma completo de 6903 verbos cuxas formas poden aparecer ben illadas ben en concorrencias con clíticos ou segunda forma do artigo.

2.2.2.2.5. *Verbos irregulares: modelos e raíces*

Polo que respecta aos verbos irregulares, o primeiro dos modelos, o Vi1, cuxo representante é SACAR, creouse pola necesidade de darlles cabida ás alternancias gráficas *c/qu*, *g/gu*, *z/c* e *u/ü* que se producen na raíz do verbo dependendo de se a vogal que segue é *a*, *o* ou *e* respectivamente:

(*sacar*) *saco, sacas / saque...*

(*achegar*) *achego, achegas / achegue...*

(*cazar*) *cazo, cazas / cace...*

(*minguar*) *minguo, minguas / mingüe...*

Debido á alteración ortográfica na raíz, para a súa implementación no lexicón necesitamos dúas raíces na conxugación sen clíticos e catro cando concorren enclíticos:

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi1 (SACAR)	Vi1a	SAC-	Vic1a	SAC-
			Vic1at	SÁC-
	Vi1b	SAQU-	Vic1b	SAQU-
			Vic1bt	SÁQU-

Este é o modelo de conjugación irregular que acolle máis verbos no seu interior cun total de 1331 lemas.

Paralelos ao modelo anterior, pero para os verbos da segunda e terceira conjugación, o Vi2 e o Vi3, cuxos representantes son COCER e TRADUCIR, acollen os verbos que alternan na raíz entre *z/c*, *g/gu* e *c/qu* dependendo de se a vogal que segue é *a*, *o* ou *e*, *i* respectivamente:

(*cocer*) *cozo, coza / coces, cocía...*

(*erguer*) *ergo, erga / ergues, erguí...*

(*verquer*⁹⁸) *verco, verca / verques, verquí...*

(*traducir*) *traduzo, traduza / traduces, traducía...*

(*cinguir*) *cingo, cinga / cingues, cinguía...*

(*delinquir*) *delinco, delinca / delinques, delinquía...*

Despois de que as *Normas* (RAG/ILG, 2003: 151-153) admitisen como normativo *ouvir* e os seus derivados, *desouvir* e *entreouvir*, implementámoslos no lexicón no grupo Vi3, aproveitando que a alternancia gráfica establecida entre as formas con *z* (*ouzo, ouza...*) e con *v* (*ouves, ouvía...*) é análoga ás recollidas no Vi3.

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi2 (COCER)	Vi2a	COC-	Vic2a	COC-
			Vic2at	CÓC-
	Vi2b	COZ-	Vic2b	COZ-
			Vic2bt	CÓZ-
Modelo Vi3 (TRADUCIR)	Vi3a	TRADUC-	Vic3a	TRADUC-
			Vic3at	TRADÚC-
	Vi3b	TRADUZ-	Vic3b	TRADUZ-
			Vic3bt	TRADÚZ-

⁹⁸ Fernández Salgado, 2004: 1216-1217: “*Variante dialectal minoritaria da forma normativa verter, que foi amplamente usada na literatura por afán diferencialista. Aparece recollida nalgúns dicionarios, pero o VOLGa só inclúe a forma practicamente xeral a todo o galego.*” En XIADA aparece na fonte deste lema *xerais_2000*.

Sen ser tan prolíficos como o Vi1 (1331), ao Vi2 remítense 217 lemas verbais e ao Vi3, 35. Con absoluto predominio da primeira conxugación, as alternancias consonánticas na raíz verbal provocadas pola vogal que segue supoñen un total de 1583 verbos, o que representa o 76,54% dos verbos irregulares galegos.

O Vi4 e Vi5, cuxos representantes son respectivamente MOER e SAÍR, créanse porque os verbos que seguen estes modelos intercalan un *i* entre a raíz e a desinencia cando esta comeza por *a* ou *o*. Este *i*, ademais, leva diérese na primeira e segunda persoas do plural do copretérito de indicativo:

(moer) moio, moia, moïamos / moes, moía...

(saír) saio, saia, saïamos / saes, saía...

Unha vez máis, o modelo clasificable na segunda conxugación, o Vi4, recibe máis lemas ca o modelo Vi5 pertencente á terceira conxugación: 30 e 6, respectivamente.

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi4 (MOER)	Vi4a	MO-	Vic4a	MO-
			Vic4at	MÓ-
	Vi4b	MOI-	Vic4b	MOI-
			Vic4bt	MÓI-
Modelo Vi5 (SAÍR)	Vi5a	SA-	Vic5a	SA-
			Vic5at	SÁ-
	Vi5b	SAÍ-	Vic5b	SAÍ-
	Vi5c	SAI-	Vic5c	SAI-
			Vic5ct	SÁI-

Os modelos Vi6 e Vi35, cuxos representantes son respectivamente AFIAR e REUNIR, xorden para dar solución na introdución ao lexicón dos verbos que manteñen o hiato na 1ª, 2ª e 3ª persoas do singular e 3ª do plural no presente de indicativo e subjuntivo e na 2ª singular do imperativo:

(afiar) afío, afías, afie / afiaba, afiei...

(acentuar) acentúo, acentúas /acentuaba, acentuei...

(reunir) reúno, reúnes, reúna / reunía, reunín...

Estes verbos, malia ser en teoría a súa conxugación totalmente regular pois seguen o modelo de CANTAR e PARTIR, na práctica non se comportan coma os regulares, xa que a alteración gráfica na raíz implica a implementación de dúas raíces no lexicón, tanto para a conxugación sen clíticos como con eles:

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi6 (AFIAR)	Vi6a	AFI-	Vic6a	AFI-
	Vi6b	AFÍ-	Vic6at	AFÍ-
Modelo Vi35 (REUNIR)	Vi35a	REUN-	Vic35a	REUN-
	Vi35b	REÚN-	Vic35at	REÚN-

Nestes dous grupos séguese mantendo a tónica xeral segundo a cal os grupos que integran verbos da primeira conxugación posúen máis elementos ca os grupos que clasifican verbos da segunda ou terceira clase. Así, no Vi6 clasifícanse 144 lemas, mentres que o Vi35 só acolle 1.

O modelo Vi30, tipo HOMOXENEIZAR, segundo o que se conxugan 16 verbos, compórtase como unha mestura do Vi1 e o Vi6, xa que comparte co primeiro a alternancia dos grafemas *z/c* e *c/qu*, mentres que co segundo participa da variación ortográfica causada porque no presente de indicativo e subxuntivo as tres primeiras persoas e a última, así como a segunda do singular do imperativo, manteñen o hiato con acento gráfico fronte ás demais formas:

(afiuzar) afiúzo, afiúzas / afiuzaba / afiúce / afiucei...

(cuincar) cuínco, cuíncas / cuincaba / cuínque / cuinquei...

Estas variacións provocan que debamos incorporar ao lexicón catro raíces, tanto na conxugación con clíticos como sen eles:

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi30 (HOMOXENEIZAR)	Vi30a	HOMOXENEIZ-	Vic30a	HOMOXENEIZ-
	Vi30b	HOMOXENEÍZ-	Vic30at	HOMOXENEÍZ-
	Vi30c	HOMOXENEIC-	Vic30b	HOMOXENEIC-
	Vi30d	HOMOXENEÍC-	Vic30bt	HOMOXENEÍC-

Os verbos da terceira conxugación rematados en *-uír*, modelo Vi7 en XIADA cuxo representante é INCLUÍR, caracterízanse por acentuar graficamente o *u* da raíz na primeira, segunda e terceira persoas do singular e terceira do plural do presente de indicativo e subxuntivo e mais a segunda persoa do singular do imperativo:

(incluír) inclúo, inclúes, inclúa...

(argüír) argúo, argúes, argúa...

A maiores, a pesar de non existir función distintiva nesta ocasión na diérese posto que o copretérito non pode confundirse co presente de subxuntivo, tamén se coloca diérese na primeira e segunda persoa do plural do copretérito de indicativo para indicar que o *i* forma sílaba de seu:

(incluír) incluïamos / incluamos

(argüír) argüïamos⁹⁹ / arguamos

Para os 49 verbos deste grupo debemos introducir no lexicón dúas raíces, unha con acento gráfico para caracterizar as formas rizotónicas e outra átona para as arrizotónicas:

Modelo Vi7 (INCLUÍR)	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
	Vi7a	INCLU- ARGÜ-	Vic7a	INCLU- ARGÜ-
Vi7b	INCLÚ- ARGÚ-	Vic7at	INCLÚ- ARGÚ-	

⁹⁹ Desviámonos un instante dos modelos formais verbais de XIADA para presentar unha cuestión que chamou a nosa atención. Trátase das dúas vogais consecutivas con diérese en *argüïamos* e *argüïades*, primeira e segunda persoas do plural do copretérito de indicativo. A diérese sobre a letra *u* emprégase, segundo Hermida (2001: 34-35), “para indicar que esta vocal se pronuncia. Isto acontece nas sílabas *güe* e *güi* fronte ás sílabas *gue*, *gui*. No primeiro caso temos que pronuncia-lo *u*, mentres que no segundo non.”

González Rei (2004: 75-76), *argúe*, con sensatez, que en galego o emprego do dígrafo *gu* se debe a un uso consagrado pola tradición, xa que non hai ningunha razón que motive a duplicidade de grafías para transcribir o fonema velar sonoro /g/. Esta dobre grafía obriga a utilizar un diacrítico máis, a diérese sobre o *u*, no caso de termos a secuencia *gue* ou *gui* nos casos en que o *u* se lea.

A nosa pregunta é por que *argüïamos* leva diérese e *argúe* non, se en ambos os casos o *u* se le? Acaso unha regra nunca explicitada impide que dous diacríticos (til e diérese) se empreguen simultaneamente? Ou débese ao feito de que por ser a vogal tónica é obvio que debe pronunciarse? Se este fose o caso, por que nos manuais de ortografía non se aclara? É máis, a pregunta que nos atrevemos a formular é por que manter un dígrafo innecesario?, que motivos impiden actualizar o noso sistema ortográfico e eliminar a grafía *gu* para o fonema oclusivo velar sonoro seguido de *e* ou *i*?

Nos verbos da terceira conjugación que teñen *e* na vogal radical debemos distinguir tres grupos:

a) O Vi8, tipo ADVERTIR. Acollerá os verbos que presentan *i* nas formas rizotónicas e *e* nas arrizotónicas, exceptuando a primeira e segunda do plural do presente de subxuntivo que leva *i* a pesar de ser a raíz átona: *advirto, advirtes, advirtamos, advirtades... / advertimos, advertides, advertín, advertía...* Segundo este modelo conxúganse 57 verbos.

b) O Vi41, tipo SERVIR. Este conxunto de 11 verbos presenta unha variación tripla no vocalismo da raíz, aínda que nós, para o seu establecemento só temos en conta a distinción entre raíces con *i* (primeira persoa singular do presente de indicativo, segunda singular do imperativo e todo o presente de subxuntivo) e raíces con *e* (todas as demais): *sirvo, sirve, sirva / serves, servimos.*

c) O Vi42, tipo SEGUIR. Os 5 verbos incluídos neste grupo caracterízanse por posuír a mesma variación vocálica ca o grupo anterior e ademais por presentar alternancia gráfica entre *g/gu*: *sigo, siga / sigue / segues, seguimos.*

As especificacións anteriores esixen que no lexicón estean representadas as distintas raíces:

	Sen clíticos		Con clíticos	
	Grupo	Raíz	Grupo	Raíz
Modelo Vi8 (ADVERTIR)	Vi8a	ADVERT-	Vic8a	ADVERT-
	Vi8b	ADVIRT-	Vic8b	ADVIRT-
			Vic8bt	ADVÍRT-
Modelo Vi41 (SERVIR)	Vi41a	SERV-	Vic41a	SERV-
			Vic41at	SÉRV-
	Vi41b	SIRV-	Vic41b	SIRV-
			Vic41bt	SÍRV-
Modelo Vi42 (SEGUIR)	Vi42a	SEGU-	Vic42a	SEGU-
			Vic42at	SÉGU-
	Vi42b	SIG-	Vic42b	SIG-
			Vic42bt	SÍG-
	Vi42c	SIGU-	Vic42ct	SÍGU-

Para os verbos da terceira conxugación que teñen *u* na vogal radical e que presentan alternancia nesta entre *u/o* temos que distinguir dous modelos distintos en XIADA:

1) O Vi9, tipo FUXIR, con *o* na segunda e terceira do singular e terceira do plural do presente de indicativo e *u* no resto da conxugación: *fluxo, fluximos, fluxe... / foxes, foxe, foxen*. De acordo con este modelo caracterízanse 32 verbos no noso lexicon.

2) O Vi29, tipo LUCIR, que mantén as características do grupo Vi9 e no que ademais as grafías *z/c* e *g/gu* están en distribución complementaria: *luzo, luza / lucimos, lucía / loces, loce; xungo, xunga / xunguimos, xunguía / xongues, xongue*. Segundo este modelo descríbense 9 verbos en XIADA.

	Sen clíticos		Con clíticos	
	Modelo Vi9 (FUXIR)	Vi9a	FUX-	Vic9a
			Vic9at	FÚX-
Vi9b		FOX-	Vic9bt	FÓX-
Modelo Vi29 (LUCIR)	Vi29a	LUC-	Vic29a	LUC-
			Vic29at	LÚC-
	Vi29b	LUZ-	Vic29b	LUZ-
			Vic29bt	LÚZ-
	Vi29c	LOC-	Vic29ct	LÓC-

Malia que nun primeiro momento poida parecer que o Vi9 e o Vi29 son o correlato do Vi41 e Vi42 por presentar alternancia vocálica nun caso e alternancia vocálica e grafemática no outro, non é exactamente así, polo que non podemos remitilos ao mesmo modelo formal. No caso da vogal radical *e* todas as raíces arrizotónicas se grafan con *e* –aínda que na pronuncia se realice como semipechada ou como semiaberta–, excepto no presente de subxuntivo onde se emprega *i*, polo que na conxugación con clíticos debemos ter unha raíz SIRV- / SIGU- para as concorrencias da primeira e segunda do plural do presente de subxuntivo con pronomes enclíticos (*sirvámoslle unha copa*). Pola contra, cando a vogal radical é *u* ningunha forma arrizotónica se grafa con *o*; non se precisa entón unha raíz átona *FOX- / *LOZ- na conxugación con clíticos, porque os tres graos de abertura da vogal radical *e* tradúcense en dous cando se trata de vogal radical *u*, e así só aparece a realización velar media, sempre ademais semiaberta, na segunda do singular e na terceira do singular e plural do

presente de indicativo, motivo polo que, calquera que sexa a combinación que elixamos, o conxunto constituído por unha destas formas verbais e mais os clíticos debe levar acento gráfico na vogal radical: *móxeslle as vacas?*, *dórme-lo neno ti?*, etc.

Nos modelos verbais que nos quedan por describir xúntase a irregularidade do verbo base con alteracións acentuais nos verbos derivados, en xeral, por transformarse algunhas unidades monosilábicas en polisilábicas: *ris/sorrís*; *cre/descré*; *pos/compós*; *di/predí*; *ten/abstén*; *ve/prevé*; *vin/prevín*; *fas/afás*; *pon/antepón*. Así do total de modelos que nos faltan, 9 débense a este factor:

VERBOS BASE		VERBOS DERIVADOS		
<i>sen clíticos</i>	<i>con clíticos</i>	<i>sen clíticos</i>	<i>con clíticos</i>	<i>nº de verbos derivados</i>
Vi40 (RIR)	Vic40	Vi10 (SORRIR)	Vic10	1
Vi11 (CRER)	Vic11	Vi31 (DESCRER)	Vic31	6
Vi12 (PÓR)	Vic12	Vi38 (ANTEPOR)	Vic38	29
Vi14 (DICIR)	Vic14	Vi36 (PREDICIR)	Vic36	5
Vi15 (TER)	Vic15	Vi33 (ABSTER)	Vic33	9
Vi19 (VER)	Vic19	Vi34 (PREVER)	Vic34	4
Vi20 (VIR)	Vic20	Vi37 (PREVIR)	Vic37	12
Vi22 (FACER)	Vic22	Vi32 (AFACER)	Vic32	8
Vi26 (POÑER)	Vic26	Vi39 (ANTEPOÑER)	Vic39	29

Os verbos CRER e LER (modelo Vi11) así coma RIR (modelo Vi40) son irregulares por resolvérense neles en crase algúns encontros vocálicos.

En RIR, malia que se produce un paralelismo nas raíces entre a conxugación da forma base e a derivada,

Sen clíticos		Con clíticos	
Vi40a	(RI-)	Vic40a	(RI-)
Vi10a	(SORRI-)	Vic10a	(SORRI-)
Vi40b	(RÍ-)	Vic40at	(RÍ-)
Vi10b	(SORRÍ-)	Vic10at	(SORRÍ-)

as terminacións e subetiquetas que as caracterizan non coinciden nun e noutro. Por exemplo, na conxugación sen clíticos, no Vi40b caracterízanse as formas verbais *río* (1ª do singular do presente de indicativo), *ría*, *rías*, *rían* (1ª, 2ª e 3ª do singular e 3ª do plural do presente de subxuntivo e copretérito de indicativo) e *rísemos*, *rísedes* (1ª e 2ª do plural do pretérito de subxuntivo), mentres que no Vi10b, ademais das formas anteriores, tamén se caracterizan *sorrís*, *sorrí* e *sorrín* (2ª e 3ª singulares do presente de

indicativo, 2ª singular do imperativo, así como a 3ª do plural do presente de indicativo e mais a 1ª singular do pretérito indicativo).

O caso de CRER (Vi11) e DESCRER (Vi31) difire do de RIR e SORRIR porque no verbo base na conxugación sen clíticos non existe unha raíz *CRÉ, imprescindible, non obstante, no derivado para poder caracterizar *descrés* (2ª singular do presente de indicativo), *descré* (3ª singular do presente de indicativo e 2ª singular do imperativo) e *descrén* (3ª plural do presente de indicativo):

Sen clíticos		Con clíticos	
Vi11a	(CRE-)	Vic11a	(CRE-)
		Vic11at	(CRÉ-)
Vi11b	(CR-)	Vic11b	(CR-)
Vi31a	(DESCRE-)	Vic31a	(DESCRE-)
Vi31c	(DESCRÉ-)	Vic31at	(DESCRÉ-)
Vi31b	(DESCR-)	Vic31b	(DESCR-)

Os modelos verbais que quedan por describir presentan tres particularidades: unha, todos eles constan dun tema de perfecto específico, distinto do de presente, segundo o que se conxugan o pretérito e antepretérito de indicativo, así como o pretérito e futuro de subxuntivo. Dúas, sobre sete deles fórmanse derivados que se diferencian do verbo base na acentuación dunhas poucas formas e, finalmente tres, os verbos base, a pesar de ser un número moi pequeno, son case todos eles verbos dun uso extraordinariamente frecuente.

Polo que respecta á implementación no lexicón de XIADA destes verbos, unicamente puidemos agrupar nun só modelo PODER e QUERER (Vi28); os demais verbos base conforman cada un deles un modelo distinto constituído por un único elemento: PÓR (Vi12), SER (Vi13), DICIR (Vi14), TER (Vi15), PRACER (Vi16), SABER (Vi17), TRAER (Vi18), VER (Vi19), VIR (Vi20), CABER (Vi21), FACER (Vi22), ESTAR (Vi23), DAR (Vi24), HABER (Vi25) POÑER (Vi26) e IR (Vi27). Pola contra, no caso dos derivados os diferentes modelos integran dende 4 verbos (Vi34, tipo PREVER) ata 29 (Vi38 e Vi39, PÓR e POÑER, respectivamente)¹⁰⁰.

A dificultade e complexidade que supuxo a introdución destes verbos no lexicón apréciase, a simple vista, no número de raíces e grupos aos que se derivan:

¹⁰⁰ Nos datos verbais, penúltimo apartado deste capítulo, recolleamos a relación completa de modelos verbais que integramos no lexicón.

	Sen clíticos	Con clíticos
Vi12 PÓR	Vi12a (PÓ-)	Vic12at (PÓ-)
	Vi12c (PO-)	Vic12a (PO-)
	Vi12b (POÑ-)	Vic12b (POÑ-)
		Vic12bt (PÓÑ-)
	Vi12d (PUÑ-)	Vic12c (PUÑ-)
		Vic12ct (PÚÑ-)
	Vi12e (PUX-)	Vic12d (PUX-)
	Vic12dt (PÚX-)	

	Sen clíticos	Con clíticos
Vi13 SER	Vi13a (SE-)	Vic13a (SE-)
	Vi13h (SÉ-)	Vic13at (SÉ-)
	Vi13b (FU-)	Vic13b (FU-)
		Vic13bt (FÚ-)
	Vi13c (FO-)	Vic13c (FO-)
		Vic13ct (FÓ-)
	Vi13d (SO-)	Vic13d (SO-)
		Vic13dt (SÓ-)
	Vi13e (ER-)	Vic13e (E-)
		Vic13et (É-)
	Vi13g (ES-)	
Vi13i (É-)		
Vi13f (SID-)		

	Sen clíticos	Con clíticos
Vi14 DICIR	Vi14a (DIC-)	Vic14a (DIC-)
	Vi14b (DIG-)	Vic14b (DIG-)
		Vic14bt (DÍG-)
	Vi14c (DIX-)	Vic14c (DIX-)
		Vic14ct (DÍX-)
	Vi14d (DI-)	Vic14d (DI-)
	Vic14dt (DÍ-)	

	Sen clíticos	Con clíticos
Vi15 TER	Vi15a (TE-)	Vic15a (TE-)
	Vi15f (TÉ-)	Vic15at (TÉ-)
	Vi15b (TEÑ-)	Vic15b (TEÑ-)
		Vic15bt (TÉÑ-)
	Vi15c (TIÑ-)	Vic15c (TIÑ-)
		Vic15ct (TÍÑ-)
	Vi15d (TIV-)	Vic15d (TIV-)
		Vic15dt (TÍV-)
Vi15e (TID-)		

	Sen clíticos	Con clíticos
Vi16 PRACER	Vi16a (PRAC-)	Vic16a (PRAC-)
		Vic16at (PRÁC-)
	Vi16b (PRAZ-)	Vic16b (PRAZ-)
		Vic16bt (PRÁZ-)
	Vi16c (PROUG-)	Vic16ct (PRÓUG-)
	Vi16d (PROUGU-)	Vic16d (PROUGU-)
	Vic16dt (PRÓUGU-)	

	Sen clíticos	Con clíticos
Vi17 SABER	Vi17a (SAB-)	Vic17a (SAB-)
		Vic17at (SÁB-)
	Vi17b (SAIB-)	Vic17b (SAIB-)
		Vic17bt (SÁIB-)
	Vi17c (SOUB-)	Vic17c (SOUB-)
		Vic17ct (SÓUB-)
	Vi17d (SEI-)	Vic17d (SEI-)
		Vic17dt (SEI-)

	Sen clíticos	Con clíticos
Vi18 TRAER	Vi18a (TRA-)	Vic18a (TRA-)
		Vic18at (TRÁ-)
	Vi18b (TROUX-)	Vic18b (TROUX-)
		Vic18bt (TRÓUX-)

	Sen clíticos	Con clíticos
Vi19 VER	Vi19a (VE-)	Vic19a (VE-)
		Vic19at (VÉ-)
	Vi19b (VEX-)	Vic19b (VEX-)
		Vic19bt (VÉX-)
	Vi19c (V-)	Vic19c (V-)

	Sen clíticos	Con clíticos
Vi20 VIR	Vi20a (VI-)	Vic20a (VI-)
		Vic20at (VÍ-)
	Vi20e (VIN-)	
	Vi20b (VE-)	Vic20b (VE-)
	Vi20c (VÉ-)	Vic20bt (VÉ-)
	Vi20f (VEN-)	
	Vi20d (VIÑ-)	Vic20c (VIÑ-)
		Vic20ct (VÍÑ-)

	Sen clíticos	Con clíticos
Vi21 CABER	Vi21a (CAB-)	Vic21a (CAB-)
		Vic21at (CÁB-)
	Vi21b (CAIB-)	Vic21b (CAIB-)
		Vic21bt (CÁIB-)
	Vi21c (COUB-)	Vic21c (COUB-)
		Vic21ct (CÓUB-)

	Sen clíticos	Con clíticos
Vi22 FACER	Vi22a (FAC-)	Vic22a (FAC-)
	Vi22b (FAG-)	Vic22b (FAG-)
		Vic22bt (FÁG-)
	Vi22c (FIX-)	Vic22c (FIX-)
		Vic22ct (FÍX-)
	Vi22d (FA-)	Vic22d (FA-)
		Vic22dt (FÁ-)
	Vi22e (FEIT-)	

	Sen clíticos	Con clíticos
Vi23 ESTAR	Vi23a (ESTA-)	Vic23a (ESTA-)
		Vic23at (ESTÁ-)
	Vi23b (EST-)	Vic23b (EST-)
	Vi23c (ESTIV-)	Vic23c (ESTIV-)
		Vic23ct (ESTÍV-)

	Sen clíticos	Con clíticos
Vi24 DAR	Vi24a (DA-)	Vic24a (DA-)
		Vic24at (DÁ-)
	Vi24c (DE-)	
	Vi24d (D-)	Vic24b (D-)
	Vi24b (DAD-)	

	Sen clíticos	Con clíticos
Vi25 HABER	Vi25a (HAB-)	Vic25a (HAB-)
	Vi25b (HA-)	Vic25b (HA-)
		Vic25bt (HÁ-)
	Vi25c (HEI-)	Vic25c (HEI-)
		Vic25ct (HÉI-)
	Vi25d (HOUB-)	Vic25d (HOUB-)
		Vic25dt (HÓUB-)

	Sen clíticos	Con clíticos
Vi26 POÑER	Vi26a (POÑ-)	Vic26a (POÑ-)
		Vic26at (PÓÑ-)
	Vi26b (PO-)	Vic26b (PO-)
		Vic26bt (PÓ-)
	Vi26c (PUX-)	Vic26c (PUX-)
		Vic26ct (PÚX-)

	Sen clíticos	Con clíticos
Vi27 IR	Vi27a (I-)	Vic27a (I-)
	Vi27d (Í-)	Vic27at (Í-)
	Vi27b (FU-)	Vic27b (FU-)
		Vic27bt (FÚ-)
	Vi27c (FO-)	Vic27c (FO-)
		Vic27ct (FÓ-)
	Vi27e (VOU)	Vic27d (V-)
	Vi27f (VA-)	

	Sen clíticos	Con clíticos
Vi28 QUERER	Vi28a (QUER-)	Vic28a (QUER-)
		Vic28at (QUÉR-)
	Vi28b (QUIX-)	Vic28b (QUIX-)
		Vic28bt (QUÍX-)
	Vi28c (QUEIR-)	Vic28c (QUEIR-)
		Vic28ct (QUÉIR-)

	Sen clíticos	Con clíticos
Vi32 AFACER	Vi32a (AFAC-)	Vic32a (AFAC-)
	Vi32b (AFAG-)	Vic32b (AFAG-)
		Vic32bt (AFAG-)
	Vi32c (AFIX-)	Vic32c (AFIX-)
		Vic32ct (AFÍX-)
	Vi32d (AFA-)	Vic32d (AFA-)
	Vi32f (AFÁ-)	Vic32dt (AFÁ-)
	Vi32e (AFEIT-)	

	Sen clíticos	Con clíticos
Vi33 ABSTER	Vi33a (ABSTE-)	Vic33a (ABSTE-)
	Vi33f (ABSTÉ-)	Vic33at (ABSTÉ-)
	Vi33b (ABSTEÑ-)	Vic33b (ABSTEÑ-)
		Vic33bt (ABSTEÑ-)
	Vi33c (ABSTIÑ-)	Vic33c (ABSTIÑ-)
		Vic33ct (ABSTIÑ-)
	Vi33d (ABSTIV-)	Vic33d (ABSTIV-)
		Vic33dt (ABSTIV-)
Vi33e (ABSTID-)		

	Sen clíticos	Con clíticos
Vi36 PREDICIR	Vi36a (PREDIC-)	Vic36a (PREDIC-)
	Vi36b (PREDIG-)	Vic36b (PREDIG-)
		Vic36bt (PREDÍG-)
	Vi36c (PREDIX-)	Vic36c (PREDIX-)
		Vic36ct (PREDÍX-)
	Vi36d (PREDI-)	Vic36d (PREDI-)
	Vi36e (PREDÍ-)	Vic36dt (PREDÍ-)

	Sen clíticos	Con clíticos
Vi37 PREVIR	Vi37a (PREVI-)	Vic37a (PREVI-)
	Vi37e (PREVÍ-)	Vic37at (PREVÍ-)
	Vi37b (PREVE-)	Vic37b (PREVE-)
	Vi37c (PREVÉ-)	Vic37bt (PREVÉ-)
	Vi37d (PREVIÑ-)	Vic37c (PREVIÑ-)
		Vic37ct (PREVIÑ-)

	Sen clíticos	Con clíticos
Vi38 ANTEPOR	Vi38a (ANTEPO-)	Vic38a (ANTEPO-)
	Vi38c (ANTEPO-)	Vic38at (ANTEPO-)
	Vi38b (ANTEPOÑ-)	Vic38b (ANTEPOÑ-)
		Vic38bt (ANTEPOÑ-)
	Vi38d (ANTEPUÑ-)	Vic38c (ANTEPUÑ-)
		Vic38ct (ANTEPUÑ-)
	Vi38e (ANTEPUX-)	Vic38d (ANTEPUX-)
	Vic38dt (ANTEPÚX-)	

	Sen clíticos	Con clíticos
Vi39 ANTEPOÑER	Vi39a (ANTEPOÑ-)	Vic39a (ANTEPOÑ-)
		Vic39at (ANTEPOÑ-)
	Vi39b (ANTEPO-)	Vic39b (ANTEPO-)
		Vic39bt (ANTEPO-)
	Vi39c (ANTEPUX-)	Vic39c (ANTEPUX-)
	Vic39ct (ANTEPÚX-)	

2.2.3. Cambios na normativa oficial do 2003: repercusións no lexicón

A edición revisada das *Normas ortográficas e morfolóxicas do idioma galego* (RAG/ILG, 2003), aprobada pola Real Academia Galega no ano 2003, introduciu unha serie de cambios con respecto á anterior edición de 1995. Cómpre, polo tanto, adaptar o lexicón de XIADA á nova normativa.

Os cambios que introduce o novo texto das *Normas* (RAG/ILG, 2003) van incidir de maneira distinta no noso lexicón. Entre eses cambios hai algúns que non teñen repercusión: o nome das letras, o acento nos interrogativos indirectos, o emprego dos signos de interrogación e admiración, etc.; outros, non obstante, como veremos deseguido, implican modificacións no lexicón para unhas poucas unidades léxicas no campo que indica a normatividade do lema e, finalmente, os menos, afectan a algún grupo de derivación. Expoñeremos primeiramente o cambio recollido na normativa e a continuación explicaremos como lle afecta ao lexicón e cal foi a solución que adoptamos.

1) Redución dos grupos cultos *-ct-* e *-cc-* a *-t-* e *-c-* respectivamente cando van tras as vogais *i* e *u*: *abducción, abdutor, adutor, aflición, aflito, conducción, conduta, conduto, condutor, conflito, constrición, constrictivo, construcción, construtivo, construtor, contradición, contrición, dedución, dedutivo, delito, destrución, destrutor, dicionario, distrito, ditado, ditadura, ditame, ditar, estrito, estrutura, flutuación, flutuar, frutífero, frutificar, frutuoso, indución, indutor, introdutor, introdutorio, lutuoso, obstrución, obstrucionismo, obstrucionista, obstrutivo, produción, produtivo, produto, produtor, reconstrución, redución, reduto, redutor, reprodución, reprodutor, restrición, restritivo, tradutor, vítima, Vítor, vitoria, vitorioso, xurisdición.*

A repercusión no lexicón deste cambio é dobre. Por unha banda provoca que as unidades léxicas anteriores co grupo *-ct-* ou *-cc-*, que antes da revisión da normativa oficial levaban no campo normatividade un valor positivo, deben ser actualizadas con respecto á indicación da normatividade e trocado o valor para negativo. Isto, ao fin e ao cabo, non implica maiores problemas, porque só afecta á raíz da unidade; co sistema que empregamos para implementar o lexicón unicamente debemos facer a modificación unha vez por cada entrada. Por outra banda provoca a incorporación no lexicón das novas unidades normativas co grupo reducido.

Non propoñemos substituír as entradas non normativas polas correspondentes normativas, porque ambas teñen que estar no lexicón de XIADA debido á súa presenza nos textos.

O cambio realizado no lexicón vémolos, por exemplo, aplicado ao adxectivo *estricto/estrito* ou ao par verbal *dictar/ditar*:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	strict	A0	G1	estricto	A	non	corga
xeral	principal	estrit	A0	G1	estrito	A	si	volga_2004
xeral	principal	dict	V	V1	dictar	V	non	corga
xeral	clíticos	dict	V	Ve1a	dictar	V	non	corga
xeral	clíticos	díct	V	Ve1at	dictar	V	non	corga
xeral	principal	dit	V	V1	ditar	V	si	volga_2004
xeral	clíticos	dit	V	Ve1a	ditar	V	si	volga_2004
xeral	clíticos	dít	V	Ve1at	ditar	V	si	volga_2004

A vocalización da consoante implosiva en *i* nas voces *reitor*, *reitorado*, *reitoral*, *reitoría* e *seita*; a ampliación da relación de voces coa solución patrimonial *-zo* e *-za* (*diferenza*, *espazo*, *graza*, *desgraza*, *licenza*, etc.); a incorporación de *liberdade* e *puberdade* ao grupo de palabras rematadas en *-dade*; o engadido de *chapeu*, *romeu* e *xubileu* á relación de palabras en *-eu*; *bacallau* e *pau* ás de *-au* ou *ímpeto* e *tribo* ás rematadas en *-o*; o cambio de *estudio* e *estudiar* (e demais familia léxica) para *estudo* e *estudar*, así como a aglutinación dalgunhas locucións (*acotío*, *amodo*, etc.), entran neste apartado, pois coinciden o problema que xeran no lexicón e a solución practicada co indicado para a redución dos grupos *-ct-* e *-cc-*.

2) Eliminación dos demostrativos neutros *esto*, *eso* e *aquelo*.

A diferenza da epígrafe anterior, este cambio non implica o incremento de novas entradas no lexicón, senón soamente a modificación na indicación da normatividade da forma, que non do lema, do valor positivo polo negativo:

lexicón	sublexicón	raíz	subetiqueta	lema	categoría	normativa da forma	normativa do lema	fonte
xeral	principal	esto	Enns	este	E	non	si	volga 2004
xeral	principal	eso	Enns	ese	E	non	si	volga 2004
xeral	principal	aquelo	Enns	aquel	E	non	si	volga 2004

O mesmo efecto provocou a supresión de *alomenos*, *anque*, *entonces*, *nembargantes* e *sen embargo*.

3) Inclusión do relativo posesivo *cuxo*, con flexión de xénero e número (*cuxa*, *cuxos*, *cuxas*).

Neste punto dáse exactamente o contrario do apartado anterior: o relativo *cuxo* e variantes, que figuraban no lexicón como non normativos aparecerán agora como normativos.

4) Admisión de *-bel* para as palabras cultas que teñen o sufixo latino *-BILIS* e, polo tanto, tamén nas palabras derivadas a partir destas: *estábel*, *estabelecer*, *estabelecemento*, etc.

No lexicón de XIADA xa constaban algúns adxectivos coa terminación en *-bel*, malia que o seu número era reducido. A raíz deste cambio na normativa, introducímolos sistematicamente e modificámoslles aos xa existentes o valor non normativo polo normativo actual. No futuro esperamos poder remitir ambas as variantes a un hiperlema, pero polo momento é simplemente unha idea, un desexo. Nestes momentos no lexicón figuran as formas duplas, tanto co sufixo *-ble* como *-bel*:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	amable	A0a	G29	amable	A	si	volga_2004
xeral	principal	amábe	A0a	G32	amábel	A	si	volga_2004
xeral	principal	establec	V	Vi2a	estabelecer	V	si	volga_2004
xeral	principal	establez	V	Vi2b	estabelecer	V	si	volga_2004
xeral	clíticos	establec	V	Vic2a	estabelecer	V	si	volga_2004
xeral	clíticos	estabélc	V	Vic2at	estabelecer	V	si	volga_2004
xeral	clíticos	establez	V	Vic2b	estabelecer	V	si	volga_2004
xeral	clíticos	estabélz	V	Vic2bt	estabelecer	V	si	volga_2004
xeral	principal	estabelec	V	Vi2a	estabelecer	V	si	volga_2004
xeral	principal	estabelez	V	Vi2b	estabelecer	V	si	volga_2004
xeral	clíticos	estabelec	V	Vic2a	estabelecer	V	si	volga_2004
xeral	clíticos	estabélc	V	Vic2at	estabelecer	V	si	volga_2004
xeral	clíticos	estabelez	V	Vic2b	estabelecer	V	si	volga_2004
xeral	clíticos	estabélz	V	Vic2bt	estabelecer	V	si	volga_2004

5) Preferencia para o sufixo tradicional *-aría*, aínda que se admite tamén a solución *-ería*.

Non repercute no lexicón máis ca para introducir as voces que aparecen no listado e que non estaban integradas aínda en XIADA.

A incorporación das formas *a respecto de*, *alén de*, *após*, *aquén de*, *até*, *até que*, *canto a*, *no entanto*, *porén* e do verbo *ouvir* intégrase tamén neste apartado.

6) Prescrición da acentuación proparoxítona na primeira e segunda persoas do plural do pretérito de subxuntivo: *andásemos, andásedes*.

Este cambio non supuxo unha alteración da estrutura do lexicón nin apenas modificou a gramática formal verbal, pero si provocou a pesada e monótona tarefa de incluír en cada un dos grupos de derivación verbais, onda as desinencias para as formas en /se/ con acentuación grave, as terminacións con acentuación esdrúxula. Naturalmente, as graves foron marcadas como non normativas e as esdrúxulas como normativas.

Cualificamos o labor de pesado porque foi preciso introducir manualmente os novos valores e modificar os xa existentes nos tres grupos regulares e nos corenta e dous irregulares que permiten o recoñecemento morfolóxico e a xeración do paradigma completo de calquera forma verbal galega. A maiores, posto que no lexicón de XIADA co campo sublexicón diferenciamos a conxugación verbal con enclíticos (*clíticos*) da conxugación sen clíticos (*principal*), realizamos tamén o cambio nos correspondentes grupos verbais regulares e irregulares que levan enclítico algún pronome ou a segunda forma do artigo.

Por exemplo, o grupo de derivación que facilita a caracterización morfolóxica e a identificación de todas as formas verbais dos verbos regulares da primeira conxugación é o V1, do sublexicón *principal*; é dicir, ningunha das formas verbais que aí se caracteriza se emprega coa segunda forma do artigo ou con clíticos. O Vc1a e o Vc1at, pola contra, permiten o recoñecemento e descrición de calquera forma verbal regular da primeira conxugación acompañada de calquera combinación de clíticos ou segunda forma do artigo (sublexicón *clíticos*). Pois ben, como vemos na seguinte táboa na que mostramos unicamente os datos relativos á primeira e segunda do plural do pretérito de subxuntivo¹⁰¹, o cambio acentual prescrito no 2003 supuxo nos grupos V1 e Vc1a a incorporación das terminacións e valores destacados en letra grosa:

¹⁰¹ As formas en *-ra* do pretérito de subxuntivo manteñen a acentuación proparoxítona a semellanza das do antepretérito, copretérito e pospretérito de indicativo. Nalgún momento teremos que acometer tamén a implementación da primeira e segunda persoas do plural dos tempos anteriores coa acentuación esdrúxula para facilitar o seu recoñecemento nos textos.

V1 (cantar) Raíz: cant-			Vc1a (cantar) Raíz: cant-		
<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>	<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>
asemos	es10p	non	asémos	es1p0	non
ásemos	es10p	si	asémo	es1p0	non
asedes	es20p	non	ásemo	es1p0	si
ásedes	es20p	si	ásemos	es1p0	si
aramos	es10p	si	asédes	es2p0	non
arades	es20p	si	aséde	es2p0	non
			ásedes	es2p0	si
			ásede	es2p0	si
			arámos	es1p0	si
			arámo	es1p0	si
			arádes	es2p0	si
			aráde	es2p0	si

7) Paso ao esquema *-án/-á* dalgunhas formas que anteriormente se acomodaban ao esquema *-án/-ana*: *afgán, alazán, alemán, barregán, bosquimán, capitán, catalán, ermitán, escribán, gardián, musulmán, rufián, sancristán, sultán e truán*.

No lexicón deben constar as características do esquema *-án/-ana* para que o etiquetador recoñeza e caracterice as formas existentes nos textos anteriores á normativa oficial actual, pero tamén temos que introducir nel as modificacións precisas para que as novas formas normativas sexan recoñecidas e analizadas.

As raíces e lemas das unidades léxicas que seguen o esquema xenérico *-án/-á*, tipo *irmán/irmá*, remítense en XIADA ao grupo de derivación G2, mentres que as unidades que na anterior normativa seguían o esquema *-án/-ana*, modelo *pillabán/pillabana*, remitíanse ao grupo de derivación G3. Imponse, polo tanto, cambiar o grupo de derivación ao que se remitían as unidades cuxa formación do feminino modificou a normativa do 2003: pasalas do G3 ao G2 e, así mesmo, incluír as terminacións *-ana* no G2 e *-á*¹⁰² no G3, cos respectivos plurais, como non normativas.

Se explicamos o problema cun exemplo, este percíbese máis claramente. A raíz *catal-* correspondente ao lema *catalán* enviábase ao grupo G3 para xerar e identificar as formas *catalán, cataláns, catalana, catalanas*. Porén, coa nova normativa, débense xerar e recoñecer tamén as formas *catalá* e *catalás*. Todas elas deben figurar no lexicón de XIADA, pois o etiquetador necesita por apareceren nos documentos que vai

¹⁰² Os cambios introducidos pola nova edición das *Normas* (RAG/ILG, 2003) non provoca a caracterización da terminación *-á* no G3 como non normativa, senón que se debe á ocasional presenza nos textos do CORGA dun feminino pexorativo en *-á*. Ex.: *Non nos podemos extrañar do alporizamento nervioso dunha muller enferma, é normal que as súas palabras sexan teimosas, que se mostre agora testalá*. [CORGA: Ú. Heinze; *O soño perdido de Elvira M.*, Xerais, 1982].

analizar, anteriores e posteriores a 2003. Se só alteramos o grupo de derivación ao que enviamos a raíz *catal-*, do G3 para o G2, unicamente solucionamos parte do problema, pois xerariamos as formas *catalá* e *catalás* pero non se identificarían *catalana* e *catalanas*. Non obstante, se combinamos o cambio de grupo co incremento das desinencias *-ana* e *-anas* no G2, por suposto caracterizadas como non normativas, conseguimos que todas as formas flexionadas do lema *catalán* se xeren e describan morfoloxicamente, e ademais no mesmo grupo.

En resumo, a acomodación ao esquema *-án/-á* de unidades que antes do 2003 seguían o esquema *-án/-ana* provocou o cambio no grupo de derivación ao que se remitían esas unidades e a modificación dos grupos G2 e G3, cuxas desinencias e valores actuais recollemos a continuación:

G2 (catal-án)			G3 (pillab-án)		
<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>	<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>
án	masculino singular	si	án	masculino singular	si
áns	masculino plural	si	áns	masculino plural	si
á	feminino singular	si	ana	feminino singular	si
ás	feminino plural	si	anas	feminino plural	si
ana	feminino singular	non	á	feminino singular	non
anas	feminino plural	non	ás	feminino plural	non

8) Todos os substantivos e adxectivos rematados en *-ón*, excepto cando esta terminación é un sufixo aumentativo ou deverbal pexorativo, forman o feminino en *-oa*: *bretón/bretoa*, *campión/campioa*, *ladrón/ladroa*, *león/leoa*, *patrón/patroa*, etc.

As actuais *Normas* (RAG/ILG, 2003) converteron en regra o que antes do 2003 era unha excepción:

Hai algunhas palabras rematadas en *-ón* que admiten un feminino en *-oa*, ó lado de *-ona*: *ladroa* (que tamén ten un feminino *ladra*), *leoa*, *patroa* ou *ladrona*, *leona*, *patrona*. [RAG/ILG, 1995: 67]

Así pois, este cambio na formación do feminino para os rematados en *-ón* supón case unha repetición do problema tratado na epígrafe anterior, só que aquí os grupos de derivación implicados son o G7 e o G8. Ao G7 remítanse as formacións aumentativas, os verbais pexorativos e calquera unidade que seguise o mesmo esquema formal, mentres que o G8 acollía a caracterización das unidades *ladrón*, *león* e *patrón* xunto cos correspondentes femininos en *-oa* e *-ona*.

A simplificación na formación do feminino para os substantivos e adxectivos rematados en *-ón* (*-ón/-oa*), sempre que este non sexa un sufixo aumentativo ou pexorativo (*-ón/-ona*), obrigounos a revisar todas as entradas que se remitían ao grupo de derivación G7 e modificar aquelas para as que se prescribira unha formación do feminino distinta, enviándoas ao grupo G8. Así mesmo, forzounos a caracterizar no G8 as desinencias *-ona/-onas* como non normativas:

G7 (lamb-ón)			G8 (campi-ón)		
<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>	<i>desinencia</i>	<i>subetiqueta</i>	<i>normativa2003</i>
ón	masculino singular	si	ón	masculino singular	si
óns	masculino plural	si	óns	masculino plural	si
ona	feminino singular	si	oa	feminino singular	si
onas	feminino plural	si	oas	feminino plural	si
			ona	feminino singular	non
			onas	feminino plural	non

9) Finalmente, quedannos só dous cambios máis incluídos polas *Normas* do 2003 (RAG/ILG, 2003) nos numerais. O primeiro deles, a eliminación dos multiplicativos *triple* e *cuádruple* apenas ten repercusión no sistema de XIADA. Por unha banda nós non os consideramos elementos da clase Numeral, senón membros das categorías Substantivo e Adxectivo, e pola outra mantémolos no lexicón, aínda que coa indicación de formas non normativas.

A deglutinación dos numerais cardinais entre *vinte* e *trinta* si reviste maior importancia, pois obrigounos a manter as formas aglutinadas coa indicación de non normativas na clase dos numerais e, a maiores, tivemos que posibilitar o recoñecemento das non aglutinadas. É dicir, *vinteún*, *vintedous*, etc. conviven no sistema de XIADA con *vinte e un*, *vinte e dous*, etc., só que en distintos habitáculos. As formas sintéticas residen no lexicón, mentres que as analíticas, como veremos no seguinte apartado, non están recollidas en ningures; a súa identificación e caracterización é posible porque no código interno do preprocesador hai unha serie de regras que os detecta.

2.3. Elementos problemáticos para un sistema computacional: as unidades multipalabra e as contraccións

2.3.1. Introducción

Unha das cuestións máis complicadas á hora de construír unha ferramenta computacional destinada a analizar morfoloxicamente unha lingua ten que ver co

tratamento de dous fenómenos complementarios: as unidades formadas por máis dun *token* e os *tokens* nos que se amalgaman varias unidades lingüísticas, ou sexa, unidades multipalabra e contraccións.

Antes de poder centrarnos no tratamento das unidades multipalabra e as contraccións, é sumamente importante que distingamos entre “palabra”, *token*, “unidade gráfica” ou “unidade ortográfica” e “unidade lingüística”, “unidade gramatical” ou “elemento gramatical”. A existencia das unidades multipalabra e as contraccións esixe que realicemos esta diferenciación, xa que non necesariamente hai correspondencia unívoca entre as unidades gráficas e as unidades lingüísticas, e ao revés.

Así, por exemplo, a expresión “*mesa redonda*” está formada por dúas palabras, entendendo por estas o segmento comprendido entre dous espazos en branco. Mais estamos ante unha unidade lingüística ou ante dúas? Só a súa inserción nun contexto nos aclarará se esas dúas unidades gráficas se corresponden cunha ou con dúas unidades lingüísticas. Vexamos as diferenzas cun par de exemplos:

Prendeu a luz e remexeu nos estantes, colleitando obxectos que ía depositando sobre a mesa redonda. [CORGA: B. Iglesias; *A historia escríbese de noite*, Galaxia, 2001]

As xornadas clausuraránse cunha mesa redonda e a presentación de conclusións. [CORGA: CG2000-11-18/12]

No primeiro caso estamos ante un moble con forma circular; polo tanto, dúas unidades gráficas que se identifican con outras tantas unidades lingüísticas. Dáse en consecuencia unha correspondencia unívoca, pois a cada elemento do primeiro conxunto (unidade gráfica) corresponde inequivocamente un elemento do segundo (unidade lingüística). No lexicón reflectímolos con entradas independentes ás que caracterizamos do seguinte xeito:

Raíz	Subetiqueta	Grupo	Lema	Normativa lema (2003)	Categoría lema	Fonte	Lexicón
mesa	Scf	G29	mesa	si	S	volga_2004	xeral
redond	A0	G1	redondo	si	A	volga_2004	xeral

Na segunda expresión, pola contra, atopámonos non co significado literal senón cun figurado: “*Conversa entre varios especialistas sobre un tema en que pode intervir o*

*público*¹⁰³. Ás dúas unidades ortográficas, *mesa* e *redonda*, correspóndelles unha única unidade gramatical, *mesa redonda*. Estamos, en consecuencia, ante unha mostra de unidade multipalabra e, como podemos observar, nestas unidades non se dá unha correspondencia unívoca entre unidade gráfica e unidade gramatical, pois en *mesa redonda* aos dous elementos do primeiro conxunto (dúas unidades gráficas) corresponde un só elemento do segundo (unha unidade lingüística). Esta circunstancia reflectímola no lexicón mediante a incorporación da unidade multipalabra nunha única entrada que debe incluír, a semellanza das unidades simples, toda a información necesaria para a súa caracterización morfolóxica:

Raíz	Subetiqueta	Grupo	Lema	Normativa lema (2003)	Categoría lema	Fonte	Lexicón
mesa redonda	Scfs		mesa redonda	si	S	corga	xeral

Fronte ás unidades gráficas simples (un *token*, unha unidade gramatical) ou ás multipalabra (varios *tokens*, unha unidade gramatical), as amálgamas caracterízanse por aglutinar nunha única unidade ortográfica varios elementos gramaticais. É o caso das palabras *nos*, *cunha* e *clausuraranse* presentes nas dúas oracións citadas máis arriba en relación con *mesa redonda*.

En *nos* e *cunha* prodúcese a contracción das preposicións *en* e *con* con, respectivamente, o artigo determinado masculino plural e o artigo indeterminado feminino singular. É dicir, sabemos que detrás das unidades gráficas *nos* e *cunha* se agochan en cada unha delas dous elementos lingüísticos pertencentes a dúas categorías gramaticais diferentes cuxas etiquetas e lemas non son compartidos en ningún dos casos. O mesmo sucede coa unidade ortográfica *clausuraranse*, coa salvidade de que aquí non existe supresión de ningún carácter, motivo polo que os dous compoñentes son recoñecibles graficamente.

Por pouco que reflexionemos dámonos conta de que a estrutura do lexicón non permite a introdución de unidades gráficas complexas nas que exista simultaneamente máis dunha unidade gramatical. As amálgamas de unidades, precisamente pola súa condición de unidades gráficas nas que están contraídos varios elementos, non poden recibir o mesmo tratamento ca as unidades simples ou multipalabra que conforman un

¹⁰³<http://www.digalego.com> [Consultado o 19/08/2010].

único elemento gramatical. Obviamente, é moi distinto que unha unidade, simple ou multipalabra, se caracterice na práctica de corpus cunha única etiqueta, malia que na teoría poidan corresponderlle varias, a que nunha soa unidade gráfica estean presentes ao tempo dous ou máis elementos gramaticais coas súas respectivas etiquetas e lemas. En consecuencia, o tratamento destes dous fenómenos complementarios vai diferir no noso sistema: as unidades multipalabra integraranse no lexicón, en tanto que as amálgamas vanse situar fóra do lexicón, aínda que non, por suposto, do sistema de XIADA.

2.3.2. Unidades multipalabra e contraccións na proposta de EAGLES

Na proposta de EAGLES recoméndase integrar as unidades multipalabra nas mesmas categorías ca as unidades simples, mentres que no caso das contraccións aconséllase tratalas como dúas categorías gramaticais distintas:

One problem which we encounter, insofar as the objects described are concerned, is how to deal with the two complementary phenomena of

1. Grammatical categories split into more than one graphical unit (multiwords or discontinuous words); and
2. Graphical units composed of more than one grammatical category (e.g. contractions).

In general, EAGLES recommends handling multiwords as belonging to a single grammatical category, and contractions as two separate grammatical categories, but the option of a different treatment is left open. [Monachini e Calzolari, 1996]

Non obstante, non se dan pautas de como levalo a cabo, pois o seu propósito non é ese:

What we propose here is the basic set of core features, derived from a detailed analysis of the major European lexicon and corpus projects; we do not aim at giving a completely worked out set of specifications ready to be implemented as such. This task is to be left at the level of the language specific development of concrete application lexicons. [Monachini e Calzolari, 1996]

Das unidades multipalabra, presentes na primeira cita deste apartado, non volvemos saber en toda a proposta de EAGLES; o termo *contraction*, pola súa parte, unicamente aparece recollido en Monachini e Calzolari (1996) no apartado referente ao tipo de elementos que van describir, *Objects described and level of description*, e na categoría Adposición, na cal o valor *fused* do atributo *Formation* indica que se está ante unha forma contracta.

A primeira aparición do termo *contracción* na proposta de EAGLES ten lugar na cita que encabeza este apartado en relación cos problemas que xorden no lexicón á hora de tratar coas unidades multipalabra e as contraccións.

A segunda aparición case presupon a existencia de contraccións unicamente entre as categorías Preposición e Artigo, de non ser por unha recomendación xeral realizada no marco da descrición do atributo “formación” concernente á Adposición:

In some languages, e.g. in Italian, prepositions may appear fused with articles in a unique graphical form. EAGLES encourages the treatment of these forms as separable, therefore having two tags corresponding to the two different categories, but also allows the use of a unique tag.

This general recommendation, i.e. to separate different categories and to encode them with different tags, is also applicable to other cases of contractions, such as negation, cliticisation, etc. [Monachini e Calzolari, 1996]

Mais ao longo da proposta non se volve falar da *negación* e, no referente á éncise, só se propón no nivel L2b, aquel no que se integran as peculiaridades lingüísticas propias dalgunhas linguas mais non comúns á maioría, un atributo *clíticos* con cuxos valores booleanos *clíticos*, *non clíticos* se codifica a presenza de clíticos nunha forma verbal. Aínda que no apartado concernente á combinación de clíticos profundaremos sobre isto, agora mesmo o que máis nos interesa da cita anterior é destacar o propósito de considerar *contracción* calquera aglutinación de elementos nunha única forma gráfica.

Do anterior despréndese, así mesmo, que toda contracción debe ser codificada discriminando nela as categorías gramaticais implicadas, polo tanto as etiquetas morfolóxicas completas correspondentes aos elementos amalgamados, o que implica que os compoñentes da contracción deben ser nalgún momento do proceso caracterizados individualmente. Esta tese, non obstante, contradise coa propia proposta de EAGLES e as aplicacións que das distintas linguas nela se recollen.

Contradí a proposta porque nesta inclúense como atributos, tanto na clase Adposición como na do Verbo, os valores que caracterizarán a existencia de contracción.

Se lembramos as recomendacións de EAGLES para o establecemento da Adposición, no atributo *Formation* establecíanse os valores *simple* e *fused* (Monachini e Calzolari, 1996):

ADPOSITION	Type	Formation	Gend	Numb	Case
EAG-L0	ADPOSITION				
	preposition	simple	m	sg	
EAG-L1	postposition	fused	f	pl	
	circumposition		n		
					nom
EAG-L2a					gen
					dat
					acc
EAG-L2b	Du 2nd-p		It c		
	Du sepadv				

É dicir, pártese da premisa de considerar as contraccións nas que están implicadas as preposicións elementos da clase Adposición.

Algo semellante ocorre co atributo *Clitic* na categoría verbal. Os valores *clt* e *nochl* marcan a presenza ou ausencia de pronomes amalgamados coas formas verbais, mais non se establecen as diferenzas entre as dúas categorías gramaticais involucradas (Monachini e Calzolari, 1996):

V	Type	Fin	Vf-M	Tns	P	N	G	Asp	Vce	Refl	MVf	AuF	Aux	Sep	Clt	Zu
L0	V E R B															
E	mai	fin	ind	pres	1	sg	m									
A	aux		sub	impf	2	pl	f									
G	mod		impr	fut	3											
			cond	past												
		no-	inf													
L		fin	part													
1			ger													
			sup													
L	s-aux							pfve	act	rfl	trs	prg				
2	cop							ifve	pas	norf	int	prf				
a											imp	pss				
												pph				
L			En ing				It c						hv	Du sep	clt	Ge inf
2			En bas										be	Du nos	nochl	Ge zu
b																

En suma, na proposta de EAGLES non se respecta o principio xeral establecido de codificar separadamente as dúas categorías afectadas.

Polo que se refire ao tratamento das contraccións de preposición mais artigo nas aplicacións recollidas por EAGLES para as distintas linguas europeas, imos dar conta aquí das correspondentes ao castelán, portugués, italiano e francés.

Tanto no modelo de EUROTRA para o castelán como no de GENELEX para o portugués prescindíuse do atributo *formación* recomendado por EAGLES e optouse por clasificar as contraccións nas que están involucradas as preposicións dun xeito peculiar: o castelán recorreu a unha categoría especial denominada *portmanteau*, mentres que o portugués as tratou como *unité morphologique agglutinée*:

In Spanish, some contractions of articles and prepositions can be found:

'a' + 'el' = 'al' 'de' + 'el' = 'del'

For analysis purposes, these contractions are treated at the morphological level in EUROTRA-ES grammars as samples of a special category *portmanteau* and are decomposed before reaching the configurational level (ECS).

(...)

This feature is not used in the Portuguese lexicon, but it is pertinent to the Portuguese language, where contraction phenomena occur. Following the GENELEX multilingual model, in the Portuguese lexicon, contraction phenomena occurring typically with preposition plus determiner are dealt with as a special kind of morphological unit (the *unité morphologique agglutinée*). [Monachini e Calzolari, 1996]

As aplicacións do italiano e o francés difiren das do castelán e portugués ao tratar as contraccións de preposición e artigo como elementos da categoría Preposición. En ambas as linguas reflíctese na etiqueta a aglutinación de preposición e artigo. Así no italiano clasifícase a contracción coa etiqueta correspondente á Preposición (E) seguida dos caracteres que identifican o xénero e número do artigo contraído (Monachini e Calzolari, 1996):

Prepositions in category 1 are often fused with following articles, forming the *preposizioni articolate*: 'agli' ('a'+gli'), 'ai' ('a'+i'), 'degli' ('di'+gli'), 'dello' ('di'+lo'). These compounds, which can be found instantiated only in a corpus, are tagged with the tag (E) plus the morphological features of the fused article.

Attribute	Value	It. example	It. tag
Formation	<i>simple</i>	di, a, da	E
	<i>fused</i>	dagli, agli	E/mp

A solución adoptada na aplicación francesa é practicamente idéntica á italiana (Monachini e Calzolari, 1996):

Attribute	Value	Fr. example	Fr. tag
Formation	<i>simple</i>	dans	PREP
	<i>fused</i>	au	PAU

Mais na proposta de EAGLES recóllese tamén para o francés outra visión; trátase da etiquetación que se lle aplicou no corpus IBM France ás preposicións e ás contraccións destas (Monachini e Calzolari, 1996):

The IBMF tagset has 2 classes of preposition:

1. PDEA for 'à' and 'de'
2. PREP for other prepositions.

and 4 classes of preposition-determiner:

1. PAU for 'au' ('à' + 'le')
2. PAUX for 'aux' ('à' + 'les')
3. PDES for 'des' ('de' + 'les')
4. PREPMS for 'du' ('de' + 'le')

Tal e como observamos, o etiquetario que se emprega no IBM francés é o máis específico, pois por unha banda diferencian dous tipos de preposicións e pola outra individualizan as catro contraccións que se dan en francés entre as preposicións *a* e *de* coas formas do artigo determinado.

Na nosa opinión, nin o castelán nin o portugués dan coa solución axeitada ante as contraccións, pois recorren a unha categoría creada *ad hoc* para tratalas. A clase Contracción, termo co que perfectamente poderíamos traducir as denominacións aplicadas en castelán e portugués, é unha creación, unha maneira de salvar o problema da integración destes elementos no lexicón, xa que a contracción en si non ten reflexo no inventario de categorías gramaticais de ningunha lingua, senón que se trata dunha convención á que se chegou para representar en certos casos a pronuncia de determinadas palabras cando van xuntas.

Malia que a solución adoptada no italiano e no francés nos gusta máis ca a do castelán ou portugués, non acaba de convencernos, pois por unha parte séguese ignorando a recomendación de EAGLES de tratar as contraccións como dúas categorías diferentes ao integralas nos lexicóns como elementos da clase Preposición e, por outra parte, as etiquetas que reciben os elementos contractos en italiano e francés diverxen das que reciben en caso de non atoparse ante contracción. Fixémonos senón no etiquetario do IBMF. Neste, a preposición *a* e *de* márcase coa etiqueta PDEA, mais nas contraccións destas dúas preposicións co artigo xa non se mantén esa etiqueta, senón que se empregan outras nas que chegan a mesturarse na cadea de etiquetaxe os trazos morfolóxicos do artigo (PREPMS) coas contraccións que se caracterizan (PAU, PAUX PDES).

2.3.3. Unidades multipalabra e contraccións noutros etiquetarios do galego

Como xa vimos no capítulo anterior, á parte do de XIADA, actualmente dispoñemos en galego e para galego de dous etiquetarios máis, deseñados ambos polo Seminario de Lingüística Informática (SLI) da Universidade de Vigo –un deles (Aguirre Moreno et al., 2001, 2002a e 2002b) aplicado ao corpus de traducións TECTRA, e o outro (Gómez Guinovart e López Fernández, 2009) aplicado ao *Corpus Técnico Anotado do Galego* (CTAG)–, polo que cremos que é interesante coñecer como se trataron neles as unidades multipalabra e as contraccións.

Polo que se refire ás unidades multipalabra, no etiquetario SLI-TECTRA só se codifican como tales as locucións adverbiais, conxuntivas e preposicionais. Non temos constancia da posible etiquetación de elementos multipalabra pertencentes a outras categorías gramaticais, pois en Aguirre Moreno et al. 2002a non se menciona sequera a problemática das unidades multipalabra ou das contraccións, e en 2002b unicamente se explicitan como unidades multipalabra as locucións prepositivas, conxuntivas e adverbiais. Ademais en Aguirre Moreno et al. 2001 déixase claro que

...ni se etiquetan los nombres propios compuestos (por ejemplo “Santa Euxenia de Ribeira”) como tales, sino que cada palabra dentro del nombre llevará su etiqueta correspondiente.
[Aguirre Moreno et al., 2001: 15]

Na etiquetación dos adverbios, conxuncións e preposicións formados por máis dunha palabra o etiquetario do SLI-TECTRA (Aguirre Moreno et al., 2001, 2002b¹⁰⁴) segue a liña proposta por Sampson (1995) para a codificación de unidades multipalabra, polo que na segmentación do texto en unidades se manteñen as palabras separadas e na análise non se proporciona unha etiqueta única ao conxunto, senón que cada unha das unidades gráficas do complexo multipalabra recibe

a etiqueta correspondente á súa categoría, seguida do signo “_”, da etiqueta correspondente á categoría da locución e dun número de dúas cifras onde o primeiro é o número de palabras que integran a locución e o segundo, o número correspondente ó lugar que ocupa a palabra dentro da locución. Por exemplo, na locución prepositiva “cara a” a “cara” asígnaselle a etiqueta composta NCFS_PREP21, mentres que a “a” lle corresponde a etiqueta “PREP_PREP22”. [Aguirre Moreno et al.; 2002b: 225]

O sistema de codificación empregado no *Corpus Técnico Anotado do Galego* (CTAG) (Gómez Guinovart e López Fernández, 2009) para as unidades multipalabra coincide co SLI-TECTRA nos elementos multipalabra considerados (adverbios, conxuncións e preposicións), mais difire totalmente deste na súa aplicación. No CTAG non hai un apartado específico para as unidades gramaticais formadas por máis dunha unidade gráfica, non obstante, na descrición esquemática do etiquetario empregando táboas exemplificativas, observamos que as locucións adverbiais, conxuntivas e prepositivas reciben a mesma etiqueta ca calquera adverbio, conxunción ou preposición simples. A única diferenza estriba en que se unen os compoñentes das unidades multipalabra co signo “_”. Véxanse senón os seguintes exemplos extraídos de Gómez Guinovart e López Fernández (2009: 64, 67)

Forma	Lema	Etiqueta	[lectura da etiqueta] ¹⁰⁵
hoxe	hoxe	R0	[Adverbio non apreciativo]
de_acordo	de_acordo	R0	[Adverbio non apreciativo]
de	de	S	[Preposición]
cara_a	cara_a	S	[Preposición]
porque	porque	CS	[Conxunción subordinante]
aínda_que	aínda_que	CS	[Conxunción subordinante]

¹⁰⁴ A única diferenza entre eles radica en que na versión do 2001 se empregaba o signo “&” para separar a etiqueta propia da unidade illada da que lle correspondía como membro dunha unidade multipalabra, mentres que na de 2002b se utiliza “_”.

¹⁰⁵ A información da última columna engadímola nós para facilitar a comprensión da táboa.

Polo que se refire ás aglutinacións de máis dunha unidade gramatical nunha única unidade gráfica, no esquema de codificación do SLI-TECTRA páttese de considerar (Aguirre Moreno et al., 2002b: 224) “*a ‘fusión’ de dúas palabras nunha mesma palabra ortográfica, propia das contraccións, dos enclíticos e da segunda forma do artigo*”, e aplícaselles a mesma solución aos tres fenómenos: asígnaselle a calquera amálgama “*unha etiqueta ‘composta’, formada pola etiqueta da primeira palabra seguida das etiquetas correspondente(s) á(s) palabra(s) ligada(s), todas elas separadas polo signo “_”*”.

Este sistema de codificación aplícase nas contraccións:

nel PREP_PPMS3 [“preposición” + “pronome persoal masculino singular terceira”]

nas formas verbais con enclíticos:

acercóuselle VIPES3_PPS3AR_PPS3AD [“verbo indicativo pretérito terceira singular” + “pronome persoal singular átono terceira reflexivo” + “pronome persoal átono singular terceira dativo”]

díxome VIPES3_PPS1A [“verbo indicativo pretérito terceira singular” + “pronome persoal átono singular primeira”]

nas formas verbais con enclíticos en amálgama:

díxomo VIPES3_PPS1A_PPMS3AA [“verbo indicativo pretérito terceira singular” + “pronome persoal átono singular primeira” + “pronome persoal átono masculino singular terceira acusativo”]

nas formas verbais con segunda forma do artigo:

bebe-lo VINF_ARDMS-2 [“verbo infinitivo” + “artigo determinado masculino singular segunda forma”]

nas formas verbais con segunda forma do artigo unida a un pronome enclítico:

gústalle-lo VIPRS3_PPP3AD_ARDMS-2 [“verbo indicativo presente terceira singular” + “pronome persoal átono plural terceira dativo” + “artigo determinado masculino singular segunda forma”]

e, finalmente, coas segundas e terceiras formas do pronome átono de acusativo:

bebelo VINF_PPMS3AA-2 [“verbo infinitivo” + “pronome persoal átono masculino singular terceira acusativo na forma segunda”]

colleuna VIPES3_PPFS3AA-3 [“verbo indicativo pretérito terceira singular” + “pronome persoal átono feminino singular terceira acusativo na forma terceira”]

En Aguirre Moreno et al. 2002b reproducíase un fragmento do TECTRA que contén unha unidade de tradución etiquetada morfosintacticamente onde podemos comprobar a aplicación da codificación dunha contracción e unha forma verbal cun pronome enclítico. Imos reproducir literalmente só a análise correspondente ás tres primeiras unidades gráficas galegas do fragmento “*Na cidade cóntase a historia da gran perla*” pois é suficiente para constatar o sistema de etiquetación empregado:

```
<tuv lang="gl"> <seg><ling> <mor lema="en" lema2="o" cat="PREP_ARDFS"/>
<ort>Na</ort> </ling> <ling> <mor lema="cidade" cat="NCFS"/> <ort>cidade</ort> </ling>
<ling> <mor lema="contar" lema2="se" cat="VIPRS3_PPS3AR"/> <ort>cóntase</ort>
</ling>
```

No sistema de anotación do CTAG equipárase tamén formalmente a codificación dos diversos casos onde se produce a unión de dúas ou máis unidades gramaticais (Gómez Guinovart e López Fernández, 2009: 68-69) "*como ocorre, por exemplo, na segunda forma do artigo, na enclise dos pronomes átonos, nas contraccións propias das preposicións, ou na contracción con artigo da conxunción comparativa ca.*"

En xeral, o método de codificación das contraccións e énclices no CTAG é o seguinte:

Se F é unha forma contracta ou enclítica formada pola unión das palabras $P1 / P2 / \dots / Pn$, sendo $L1, L2 \dots Ln$ os lemas das palabras compoñentes e $C1, C2 \dots Cn$ as súas etiquetas categoriais, a forma codificada xenérica da forma contracta sería $F/L1_C1 \sim/L2_C2 \dots \sim/Ln_Cn$, como se ilustra a seguir na etiquetaxe das formas contractas *facelas*, *nesoutra* e *entrámbolos*:

• *facelas/facer_VN0000* \sim/o_P3FPA0

• *nesoutra/en_S* $\sim/ese_DFS \sim/outro_IFS0$

• *entrámbolos/entre_S* $\sim/ambos_IMP0 \sim/o_GMP$ [Gómez Guinovart e López Fernández, 2009: 68]

Neste sistema non se realiza segmentación formal das unidades gráficas contractas, senón que estas considéranse un todo que é analizado desagregando o *token*

en tantos lemas e etiquetas como elementos gramaticais posúa a unidade ortográfica. Para substituír a forma correspondente a cada elemento da unidade contracta, excepto o primeiro, empregan o til de nasalidade que, segundo Gómez Guinovart e López Fernández (2009: 68), “indica que a forma fónica da palabra está subsumida na *contracción anterior*”.

A diferenza do SLI-TECTRA, no SLI-CTAG non existen as etiquetas compostas, cada compoñente gramatical da unidade gráfica recibe a mesma etiqueta e lema que lle correspondería de non atoparse nunha amálgama. O mesmo método aplícase en todos os casos de aglutinación de unidades gramaticais. Así sucede nas contraccións:

coa/con_S ~/o_GFS [“preposición”/ “artigo feminino singular”]

cara á/cara a_S ~/o_GFS [“preposición” / “artigo feminino singular”]

cós/ca_CS ~/o_GMP [“conxunción subordinante” / “artigo masculino plural”]

nas formas verbais con pronomes enclíticos:

permítenlles/permitir_VIP3P0 ~/lle_P30PD0 [“verbo indicativo presente terceira plural xénero non especificado” / “pronome terceira xénero non especificado plural dativo cortesía non especificada”]

ante as segundas formas do artigo:

tódolos/todo_IMP0 ~/o_GMP [“indefinido masculino plural grao non especificado” / “artigo masculino plural”]

collémo-la/coller_VIP1P00¹⁰⁶ ~/o_GFS [“verbo indicativo presente primeira plural xénero non especificado grao non especificado” / “artigo feminino singular”]

ou na unión dos pronomes en dativo co acusativo de terceira persoa:

cho/che_P2CSD0 ~/o_P3MSA0 [“pronome segunda común singular dativo cortesía non especificada” / “pronome terceira masculino singular acusativo cortesía non especificada”]

nola/nos_P10PC0 ~/o_P3FSA0 [“pronome primeira ¿xénero non especificado?¹⁰⁷ plural acusativo/dativo/reflexivo cortesía non especificada” / “pronome terceira feminino singular acusativo cortesía non especificada”]

¹⁰⁶ Desenvolvemos as etiquetas do CTAG para que sexan comprensibles, mais nós non eliminamos nin engadimos valores, polo que non corriximos as diferenzas na cadea de etiquetaxe entre “permítenlles” e “collémo-la”, presentes en Gómez Guinovart e López Fernández (2009: 68).

En Gómez Guinovart e López Fernández (2009: 69-70) reproducense algúns fragmentos ilustrativos tirados do CTAG onde podemos comprobar a aplicación da codificación de varias contraccións. Imos reproducir literalmente só a análise pertencente ás primeiras unidades do primeiro segmento, pois é suficiente para reflectir a análise real de corpus:

```
<frase>^A/O_GFS      ^expansión/expansión_NCFS0      ^dos/de_S      ^~/o_GMP
^cultivos/cultivo_NCMP0  ^transxénicos/transxénico_A0MP  ^ameaza/ameazar_VIP3S00
^a/o_GFS ^diversidade/diversidade_NCFS0 ^xenética/xenético_A0FS ^pola/por_S ^~/o_GFS
^simplificación/simplificación_NCFS0 ^dos/de_S ^~/o_GMP ^sistemas/sistema_NCMP0 .....
```

Sobre este fragmento analizado interézanos especialmente destacar dous detalles; o primeiro deles ten que ver coa representación en maiúsculas do primeiro lema da frase correspondente ao artigo determinado fronte á minúscula empregada para o mesmo lema no interior da frase:

```
^A/O_GFS ^expansión/expansión_NCFS0 fronte a ^a/o_GFS ^diversidade/diversidade
```

Isto implica que todas as unidades que encabezan unha frase presentan un lema distinto do desas mesmas unidades se se atopan no interior da frase? Non sabemos a resposta xa que en Gómez Guinovart e López Fernández (2009) non se fai ningunha referencia a iso.

En segundo lugar, vémonos obrigados a corrixir unha observación anterior nosa realizada baseándonos no método de descrición achegado polos autores. Refírese á ausencia de segmentación nas unidades contractas, pois, se nos fixamos no fragmento anotado, o signo “^” marca o comezo da análise de cada unha das unidades gramaticais: *^A/O_GFS*, *^expansión/expansión_NCFS0*, etc. Ese mesmo signo precede lóxicamente ás formas contractas *dos* e *pola*, mais tamén precede á análise do segundo compoñente coma se este aparecese illado no texto, coa salvidade de que a forma gráfica que debería aparecer se ese fose o caso substitúese co signo “~”:

```
^dos/de_S ^~/o_GMP [“preposición”/ “artigo masculino plural”]
```

```
^pola/por_S ^~/o_GFS [“preposición”/ “artigo feminino singular”]
```

¹⁰⁷ Segundo a táboa de atributos e valores e os exemplos (Gómez Guinovart e López Fernández, 2009: 65) nesta etiqueta para expresar o valor do xénero debería aparecer “C” (común), mais o que en realidade figura é “0”.

Isto bate coa descrición xeral do método de codificación das contraccións na cal se afirma que a forma codificada xenérica da forma contracta é $F/L1_C1 \sim/L2_C2... \sim/Ln_Cn$, de onde nós deduciamos que $F/L1_C1 \sim/L2_C2... \sim/Ln_Cn$ constitúe unha única unidade, aínda que despois se desagreguen os lemas e etiquetas de cada un dos compoñentes. É este o motivo polo que o formato escollido para reflectir a anotación morfolóxica das formas aglutinadas, na nosa opinión, non se axusta ao obxectivo teórico pretendido.

O sistema de codificación escollido polo SLI-TECTRA para anotar as unidades multipalabra e as contraccións, por outra banda, tal e como se afirma en Aguirre Moreno et al. (2001: 15) presenta a tara de aumentar o número de etiquetas do repertorio, debido precisamente á existencia das etiquetas compostas, en comparación xa non cos etiquetarios doutras linguas con moita menor presenza destes elementos, senón tamén con outros etiquetarios do galego como son o do CTAG ou o de XIADA, onde tamén se codifican as unidades multipalabra e os diferentes tipos de amálgamas.

2.3.4. Unidades multipalabra e contraccións en XIADA

2.3.4.1. Unidades multipalabra

A priori, o galego non se diferencia *grosso modo* doutras linguas románicas no que respecta á existencia e tipoloxía de unidades gramaticais constituídas por máis dunha unidade gráfica; así os adverbios, preposicións e conxuncións, compostos por varias unidades ortográficas, trátanse morfoloxicamente no sistema de XIADA coma se estiveramos ante unha unidade simple, tal e como vimos no apartado 2.2.2.1., correspondente á implementación no lexicón do vocabulario clasificable categorialmente en clases de palabras invariables. Dado que as unidades multipalabra conforman unha porción do corpus léxico galego, deben incorporarse ao lexicón, mais tendo en conta loxicamente as súas peculiaridades formais.

Debemos aclarar, con todo, que este é un capítulo non pechado aínda, pois malia que o groso das unidades multipalabra pertencentes ás categorías invariables (adverbios, conxuncións e preposicións) si se introduciron xa no lexicón, non sucede o mesmo coas unidades multipalabra que presentan diferentes modos de flexionar o xénero e o número. Isto é debido, entre outros factores, á nosa pretensión de atopar unha maneira

de formalizar as diferentes estruturas internas destas unidades para implementalas no lexicón de xeito que non teñamos que introducir todas as combinacións posibles en cada un dos casos.

Esta é a causa que explica que, polo momento, o número de unidades multipalabra pertencentes a categorías variables que figuran no lexicón sexa moi reducido (*alto cargo, best seller, disco duro, garda civil, mass media, mesa redonda, primeiro ministro, prime time e visto e prace*), pois preferimos dilatar un pouco máis a introdución deste tipo de elementos á espera de ser capaces de implementalos de xeito análogo aos demais lemas que integran o lexicón.

Unha posible vía de acción consistiría en identificar os distintos compoñentes das unidades multipalabra, de xeito que cada un deles estivese en relación coa correspondente unidade simple e a caracterización de xénero e número desta, mais ao tempo teriamos que poder indicar trazos xenéricos e numéricos distintos por se a flexión da unidade multipalabra difire da da suma dos seus compoñentes. É dicir, unha unidade como *primeiro ministro* poderíamos desagregala en dous compoñentes: *primeiro* e *ministro*. Logo vendo que a flexión da unidade multipalabra (*primeiro ministro, primeira ministra, primeiros ministros, primeiras ministras*) se corresponde coa flexión dos seus compoñentes, considerados estes como unidades simples, poderíamos interconectar computacionalmente as tres unidades, as dúas simples e mais a multipalabra, de xeito que compartisen a información relativa a raíces e grupos de derivación. Así, a información que posuímos no lexicón sobre os dous compoñentes da unidade multipalabra escollida é a pertencente á súa caracterización como elementos lingüísticos autónomos:

Raíz	Subetiqueta	Grupo	Lema	Normativa	Categoría	Fonte	Lexicón
primeir	Noa	G1	primeiro	si	N	volga_2004	xeral
ministr	Sc	G1	ministro	si	S	volga_2004	xeral

Pois ben, se conseguíramos introducir a raíz *primeir ministr* cuxo lema sería *primeiro ministro* e remitir cada un dos compoñentes ao grupo de derivación G1, sen perder de vista a unidade multipalabra, estaríamos acadando parte do noso obxectivo: elaborar modelos formais que posibiliten a incorporación de unidades multipalabra ao lexicón. Non obstante, a solución adoptada para o lema *primeiro ministro*, válida tamén

para casos coma *primo irmán* (*prima irmá / primos irmáns / primas irmás*) ou *garda civil* (*gardas civís*) non sería axeitada para unidades coma *ano luz* (*anos luz*) ou *orella de mar* (*orellas de mar*), pois a flexión de xénero e número da unidade multipalabra non se adecúa á dos seus compoñentes. É preciso, en consecuencia, profundar no estudo destas unidades e, en coordinación co enxeñeiro informático, decidir cal é a mellor opción para a súa incorporación ao lexicón.

2.3.4.2. Contraccións

De todos é coñecido que en galego as contraccións que teñen lugar na lingua escrita presentan unha diverxencia considerable, por exemplo, respecto do castelán, e en menor medida do portugués, sobre todo debido ao seu grande número, mais tamén pola amplitude de categorías que nelas están implicadas. Non obstante, á hora de determinar como se van tratar non importa que esteamos ante un número reducido ou amplo, que haxa unha ou máis categorías gramaticais implicadas, que sexan dous ou máis de dous os elementos aglutinados... Posto que o termo contracción implica a aglutinación de máis dun elemento gramatical cos seus correspondentes lemas e etiquetas, e posto que no lexicón non hai cabida para estas unidades, temos que decidir en que lugar, de que maneira e en que momento do proceso de recoñecemento imos caracterizar a unidade contraída.

Como conxugar a información pertinente para a caracterización morfolóxica dos *tokens* amalgamados –unidade gráfica, unidades gramaticais e categoría, etiqueta e lema de cada unidade gramatical– co seu tratamento computacional vai centrar o contido das seguintes epígrafes.

Comezaremos por presentar a homoxeneidade que nas diferentes obras de referencia da gramática galega atopamos á hora de reflectir a heteroxeneidade das distintas unidades aglutinadas.

É curioso constatar como un mesmo feito, a aglutinación de distintas unidades gramaticais nunha única forma gráfica, recibe diferente denominación dependendo do tipo de aglutinación que teña lugar. Así, o normal é que o encontro de preposición con artigo determinado ou indeterminado, demostrativo, pronome persoal tónico, indefinido... reciba o nome de *contracción* (RAG/ILG, 1995: 87, 89, 92, 98, 100 e 102;

RAG/ILG, 2003: 85, 88, 91, 97, 99 e 101; Álvarez, Regueira e Monteagudo, 1986: 139, 141, 162, 208, 234, 236 e 245; Álvarez e Xove, 2002: 448-449, 457-458, 468, 486, 509 e 516; Freixeiro, 2000: 122, 199, 251, 270 e 593); o encontro entre os pronomes de dativo co acusativo de terceira persoa é denominado maioritariamente *concorrenza* (RAG/ILG, 1995: 93-94; RAG/ILG, 2003: 92-93; Álvarez, Regueira e Monteagudo, 1986: 165-166), aínda que outros autores prefiren consideralo *amálgama* (Álvarez e Xove, 2002: 539) ou mesmo *contracción* (Freixeiro, 2000: 124); finalmente, a unión do artigo determinado coas preposicións *por* e *tras*, coas formas verbais rematadas en *-r* ou *-s*, etc. non é considerada *contracción*, senón que en xeral fálase de alomorfos ou variantes en distribución complementaria e da súa representación gráfica (RAG/ILG, 1995: 84-87; RAG/ILG, 2003: 83-85; Álvarez, Regueira e Monteagudo, 1986: 136-138; Álvarez e Xove, 2002: 443-447), aínda que tamén en ocasións se describen estes casos como *asimilación* (Freixeiro, 2000: 261-262), ou incluso *amálgama* se non hai trazo separador (Álvarez e Xove, 2002: 443).

De acordo co DRAG (1997) no ámbito gramatical unha *contracción* é a “unión de dúas palabras pola perda dunha vocal ou consoante nunha delas” mentres que á *amálgama* non se lle recoñece unha acepción gramatical e só é definida como “alixe” na primeira acepción e como “mestura” na segunda. Tampouco para *combinación* ou *concorrenza* se recolle neste dicionario unha acepción lingüística específica.

No dicionario de Ir Indo en liña¹⁰⁸ a *contracción* é definida no terreo lingüístico como a “Redución dun elemento lingüístico, a miúdo en combinación con outro elemento” e “Palabra que resulta desta redución”, en tanto *amálgama* consiste na “Fusión, nun morfema, de dous significados que adoitan representarse con dous significantes diferentes”.

No Xerais (Carballeira Anllo, 2009) a *contracción* defínese como o “Fenómeno de fusión de dúas vogais en contacto nunha soa (p.ex.: *ao* > *ó*)” e tamén como a “[Impropiamente] *amálgama* ou *asimilación* de dúas vogais nunha”, mentres que a *amálgama* consiste na “Fusión de dúas palabras ou de dous morfemas nun só”.

En Hermida (2001) contamos cun apartado específico no que polo seu encabezamento –*V. A ortografía da palabra. Composición, contracción, asimilación e*

¹⁰⁸ <http://www.digalego.com> [consultado o 18/8/2010]

amálgama– esperamos atopar claridade e precisión na definición e emprego destes termos, mais non se cumpre o noso desexo, pois malia que a autora establece unha distinción entre *contracción* e *asimilación*

As normas ortográficas do galego optaron por representar na escrita as palabras illadas e na súa plenitude, aínda que nalgúns casos consagraron determinadas asimilacións (unión de dúas palabras coa perda da vocal dalgunha delas) e contraccións (unión de dúas vocais co resultado dunha vocal distinta destas dúas) propias da fala. [Hermida, 2001: 50]

logo non a aplica, e así recolle en cadros as “Contraccións e asimilacións do artigo determinado” (2001: 50), as “Asimilacións do artigo indeterminado” (2001: 51), as “Asimilacións do demostrativo” (2001: 51), as “Asimilacións da combinación do demostrativo co indefinido *outro*” (2001: 52), as “Asimilacións do pronome tónico” (2001: 52) e a “Concorrenza de varios pronomes átonos” para o encontro dos pronomes átonos de dativo co de acusativo de terceira persoa (2001: 52), mentres que nas breves descrições emprega, indistintamente na nosa opinión, variantes dos termos *amálgama* e *contracción*.

Finalmente decidimos acudir aos dicionarios de termos lingüísticos para ver se nestes atopabamos a solución, mais tampouco nos resultaron demasiado útiles. En Abraham (1981: 51, 56) non se recollen as voces *contracción* e *concorrenza*, e as definicións de *aglutinación* e *amalgama* non nos parecen moi claras; Mounin (1979: 9, 13-14, 37, 46) e Dubois et al. (1992: 24-25, 38, 111, 148) coinciden *grosso modo* en *aglutinación*, ao tratala como erros de segmentación (*el arradio < la radio*), e en *amalgama* (dous significados normalmente representados por dous significantes distintos confúndense nun só significante inanalizable), mentres que a *contracción* ou *coalescencia* é definida como a fusión de dous ou máis elementos fónicos nun só; mais, curiosamente, Mounin inclúe as formas *al* e *del* como exemplos casteláns tanto de *amalgama* como de *contracción*, en tanto Dubois et al. os consideran só *amalgama*, xa que con *coalescencia* parecen referirse máis ben á crase vocálica.

Como acabamos de comprobar, a diferenza entre os distintos termos empregados para denominar a aglutinación de varias unidades lingüísticas nunha soa unidade gráfica non está demasiado clara; non obstante, non imos ser nós quen a establezamos aquí, xa que dende o punto de vista computacional non é necesario debido a que o tratamento destas unidades ten que ser semellante. Por iso, a partir deste momento non faremos

distincións entre contracción, asimilación, amálgama, combinación, etc., e calquera destes termos poderá ser empregado para referenciar o mesmo concepto: varias unidades gramaticais nunha única unidade ortográfica.

Si imos establecer, con todo, unha diferenza neste *totum revolutum* á hora de implementalo no sistema de XIADA, mais esta virá determinada pola pertenza da unidade gráfica a un grupo limitado de elementos (contraccións) ou a un grupo aberto (forma verbal con enclíticos ou segunda forma do artigo). É dicir, malia que todas as unidades gráficas nas que se produce aglutinación de elementos lingüísticos van ser desagregadas nos seus constituíntes, caracterizando morfoloxicamente cada un deles, a información relativa ás contraccións xestionarase dun xeito distinto á información relativa ás formas verbais con segunda forma do artigo ou formas verbais con calquera combinación de pronomes enclíticos, incluídas as amálgamas dos pronomes de dativo cos de acusativo de terceira. A explicación é obvia; podemos introducir nunha base de datos todas as contraccións; será un proceso custoso ou monótono, pero é factible, posto que nos enfrontamos cun número limitado de unidades, 203 en concreto. Pero podemos facer isto mesmo ou, mellor dito, é aconsellable desenvolver o mesmo proceso con cada unha das formas verbais de cada un dos verbos que figuran no lexicón que se poden construír coa segunda forma do artigo ou ás que se lles poden engadir ata tres-catro pronomes enclíticos? Naturalmente, o sentido común indícanos que non é ese o camiño a seguir, senón que debemos buscar un atallo efectivo que nos permita partir do mesmo punto, a unidade gráfica, para chegar ao mesmo destino, o recoñecemento e caracterización da forma verbal e mais de cada un dos pronomes enclíticos ou segunda forma do artigo. Deseguido veremos como se integran as contraccións en XIADA e na epígrafe seguinte describiremos a implementación das unidades integradas por unha forma verbal e un ou varios elementos enclíticos.

En primeiro lugar centrarémonos sobre como imos caracterizar as unidades contractas, para logo ver onde as imos colocar e, finalmente, describir cando e como o sistema de recoñecemento se enfronta a elas.

O paso primeiro consiste en elaborar un inventario de todas as contraccións posibles que se representan na lingua escrita. Para iso, recorreremos a Álvarez, Regueira e Monteagudo (1986) e ás *Normas* (RAG/ILG, 1995). Logo trasladamos esa información

a un formato tabular no que cada unha das amálgamas ocupa unha fila diferente, e nas columnas indicamos a forma, lema e etiqueta do primeiro dos constituíntes; a continuación o signo “/”, único elemento da columna correspondente, exerce de separador da caracterización dos compoñentes da contracción e, finalmente, nas tres últimas columnas da mesma fila figuran a forma, o lema e a etiqueta do segundo dos compoñentes da contracción. Esta descrición coincide coa maioría das contraccións incluídas no sistema de XIADA e, en concreto, por exemplo, coas amálgamas do artigo determinado e mais as preposicións *en* ou *de*:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
do	de	de	P	/	o	o	Ddms
dos	de	de	P	/	os	o	Ddmp
da	de	de	P	/	a	o	Ddfs
das	de	de	P	/	as	o	Ddfp
no	en	en	P	/	o	o	Ddms
nos	en	en	P	/	os	o	Ddmp
na	en	en	P	/	a	o	Ddfs
nas	en	en	P	/	as	o	Ddfp

A non distinción no noso sistema entre contracción e aglutinación permítenos tratar do mesmo xeito todas as combinacións da preposición *a* co artigo determinado, pois se nas formas *ó*, *ós*, *á* e *ás* nos atopamos en realidade cunha contracción, non sucede o mesmo coas formas *ao* e *aos*, dado que nestas últimas a contracción prodúcese na fala mais non na escrita, onde só ten lugar a aglutinación. En XIADA a caracterización de *ó(s)* e *ao(s)* é idéntica:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
ó	a	a	P	/	o	o	Ddms
ós	a	a	P	/	os	o	Ddmp
ao	a	a	P	/	o	o	Ddms
aos	a	a	P	/	os	o	Ddmp
á	a	a	P	/	a	o	Ddfs
ás	a	a	P	/	as	o	Ddfp

Malia que na proposta de EAGLES e nas definicións en xeral do termo contracción se acostuma falar do encontro só de dúas palabras, en galego atopámonos nun par de casos con que é necesario darlle cabida a un terceiro constituínte, pois a amálgama consta de tres compoñentes. É o que ocorre cando *estoutro*, *esoutro* e *aqueloutro* –e variantes– contraen coas preposicións *de* e *en* ou cando se utiliza

incorrectamente a unidade gráfica non indefinida *entrambos*, *entrambas* coa segunda forma do artigo.

Nos últimos tempos a estrutura e emprego da unidade *entrambos* deveu en controvertible pois, ante a falta dunha gramática oficial e mais a ausencia dunha explicación sobre o seu uso nas *Normas* (RAG/ILG, 1995, 2003), non está claro se *entrambos* se pode utilizar coma unha suma de ‘entre+ambos’ ou, máis ben, coma unha alternativa sinónima de *ambos*.

As autoridades lexicográficas do galego non nos axudan a resolver o problema: o DRAG (1997, 2012) considera *entrambos-as* a contracción de *entre* e mais *ambos-as* (*Tiña dous fillos e coidábanos entrambos*), mentres que no VOLGa (González e Santamarina, 2004) o indefinido é tratado como unha única unidade lingüística. Compárense, senón, as entradas correspondentes a *entrambos* e *entrámbolos* (cfr. 349):

entrambos *p f*: **entrambas**

entrámbolos *cont f*: **entrámbalas**

Nota: *entrambos(p)+los(art)*

Polo que se refire ás gramáticas, Álvarez, Regueira e Monteagudo, pola súa banda, afirman:

Entrambos, *entrambas*, procede de *entre* + *ambos/ambas* e, en principio, mantén o significado de 'un e mailo outro, entre os dous' (*fixérono entrámbolos dous; coidárono entrámbalas irmás*); con todo, hoxe úsase como sinónimo de *ambos*, presentando as mesmas características:

Tanto ten unha garavata coma a outra: *entrambas* son iguais.

Vanche *entrámbalas dúas* na festa.

Ó final, *entrámbolos* nenos acabaron por chorar. [Álvarez, Regueira e Monteagudo, 1986: 262]

Freixeiro (2000: 257) constata, sen máis, a súa existencia, e en Hermida (2004: 48-49) só podemos deducir pola descrición que o considera unha variante de *ambos*.

Temos que acudir a Álvarez e Xove (2002: 490) para obter unha resposta clara:

A variante xeral é *amb-*, coa que alterna libremente o alomorfo *entramb-*. [...]

Entrambos non é *ambos* precedido da preposición *entre*; esta preposición, coma outras, pode preceder a *ambos* e a *entrambos*:

O xarapón xa o tiveron *ámbalas irmás* ~ *entrámbalas irmás*. Podes xogar con *ambos* ~ con *entrambos*.

Entre ámbolos ~ Entre entrámbolos xemelgos só hai pequenas diferencias¹⁰⁹.

Así pois, *ambos* e *entrambos* son dúas variantes que expresan o mesmo significado: *os dous, un e mais o outro*.

En XIADA, non obstante, tendo en conta os textos do CORGA, cremos que estamos ante dúas unidades distintas: unha correspondente ao indefinido *entrambos* e outra pertencente á contracción incorrecta da preposición *entre* co indefinido *ambos*. Nos seguintes exemplos, na nosa opinión, os tres primeiros casos representan o indefinido, mentres que os tres últimos son erros por *entre ambos* e exemplifican a contracción:

"vin tamén a Sísifo, quen padecía duros traballos, empuxando con entrambas mans unha enorme pedra; [CORGA: F. M. Antón e M. Mandianes; *O ciclo da vida*, Ir Indo, 1998]

Deitados nun xergón de folla de millo, botaban entrambos as horas a palicar mentres o avó, coa man, lle agarimaba as costas. [CORGA: X. M. Valcárcel, *Os ollos da sentinela*, Galaxia, 1997]

A forma xurídica máis empregada por este tipo de empresas é a sociedade limitada (ver gráfica 1) seguida da sociedade anónima, englobando entrambas preto do 80%. [CORGA: P. Vázquez Rodríguez, A. Gayo López e M. A. López Cabarcos, *Revista Galega de Economía: A situación actual e de futuro dun viño novo: a denominación de orixe Rías Baixas*, Universidade de Santiago de Compostela, 2001]

Touriño asinou onte un protocolo para fomentar a rodaxe de coproducións entrambos os países e mellorar a súa promoción [CORGA: *GH2007-03-17/47*]

Publicado no ano en que seu pai é protagonista, o libro foi presentado xunto con Canle Segredo, de Xosé María Álvarez Blázquez, tamén na colección Dombate de Galaxia, case como un diálogo entrambos os dous. [CORGA: *ANT2008-12-04/82*]

"Este feito _comenta o autor_ xustifica tanto o interese pola relación e coincidencias biopoéticas existentes entrambos autores coma o seguimento da débeda lírica que estes contraeron co barco insignia do vangardismo galaico: Manuel Antonio". [CORGA: *CG2000-12-03/41*]

Cando, coas dúas unidades anteriores, se emprega a segunda forma do artigo atopámonos sempre ante dúas contraccións; non obstante, a súa segmentación e caracterización variará, pois nunhas ocasións posuirá dous compoñentes e noutros tres. Así, ante os exemplos seguintes

¹⁰⁹ Ante este estraño exemplo decidimos comprobar se os corpus galegos actuais avalaban este uso, pero na consulta realizada o 7 de setembro do 2010 sobre o CORGA na súa versión 1.5 en probas (<http://corpus.cirp.es/corgadev>; 111 ocorrencias), o TILG (<http://www.ti.usc.es/tilg/index.asp>; 557 exemplos), o CTAG (<http://sli.uvigo.es/CTAG/index.html>; 5 ocorrencias) e o CTG (<http://sli.uvigo.es/CTG/index.html>; 12 casos) en ningún dos resultados estaba precedido *entrambos* da preposición *entre*.

Deduzo isto polo asombro que se nota nos rostros que tendes colocados xusto no medio de entrámbolos ombreiros. [CORGA: A. Cortizas Amado; *O coleccionista de sombras (Farsa en tres actos con dous entre actos)*, Sotelo Blanco, 1990]

Por moita amizade que exista entrámbalas partes, non está en terreo conquistado e a súa educación tanto como o respecto que lle merecen os visitados, ten que manifestalos. [CORGA: L. Prego Carregal; *Saber ser, saber estar. Normas-Consellos-Curiosidades*, Toxosoutos, 1994]

corresponde aplicar análises distintas, xa que no primeiro caso, en *entrámbolos ombreiros*, temos que caracterizar dous constituíntes, o indefinido *entrambos* e mais o artigo:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
entrámbolos	entrambos	entrambos	Idmp	/	los	o	Ddmp
entrámbalas	entrambas	entrambos	Idfp	/	las	o	Ddfp

e no segundo, en *entrámbalas partes*, son tres os constituíntes, pois ademais do compoñente pertencente ao artigo, *entrambos* ten que ser desagregada en dous elementos lingüísticos, a preposición e o indefinido:

	Constituínte 1				Constituínte 2				Constituínte 3		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
entrámbolos	entre	entre	P	/	ambos	ambos	Idmp	/	los	o	Ddmp
entrámbalas	entre	entre	P	/	ambas	ambos	Idfp	/	las	o	Ddfp

Cuestionable é tamén nos últimos anos a aglutinación do demostrativo masculino e feminino co indefinido *outro* en *estoutro*, *esoutro*, etc. O DRAG (1997: 103, 495 e 511) defíneas como unidades demostrativas que poden contraer coas preposicións *de* e *en*; as *Normas* (RAG/ILG, 1995: 97; 2003: 96) trátaas como “combinacións do demostrativo co indefinido *outro*”, en tanto no VOLGa observamos unha incoherencia, xa que a consideración de *aqueloutro* como contracción difire da de pronome de *estoutro* e *esoutro* (2004: 106, 366 e 376):

aqueloutro *cont* f: **aqueloutra**

Nota: aquel(p)+outro(p)

esoutro *p* f: **esoutra**

Nota: ese(p)+outro(p)

estoutro *p* f: **estoutra**

Nota: este(p)+outro(p)

En XIADA optamos por considerar as unidades *estoutro*, *esoutro* e *aqueloutro* –coas súas correspondentes formas flexionadas– cando se amalgaman con *de* e en contraccións formadas por tres constituíntes, posto que, se consideráramos estas unidades unha única unidade lingüística demostrativa, deberíamos ser consecuentes e tratar do mesmo xeito, malia que como unidades multipalabra, as ocorrencias analíticas de demostrativo seguido de *outro*, pero cremos que non procede esa análise. Computacionalmente non nos supón ningún problema, pois só temos que continuar coa serie previamente establecida para as contraccións de dous elementos: un novo separador “/” seguido da forma, lema e etiqueta do novo compoñente. Vexamos, por exemplo, o contraste entre as contraccións do demostrativo con *outro* –formadas por dous constituíntes– e as destas coa preposición *de* –integradas por tres constituíntes–:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
estoutro	este	este	Edms	/	outro	outro	Iams
estoutros	estes	este	Edmp	/	outros	outro	Iamp
estoutra	esta	este	Edfs	/	outra	outro	Iafs
estoutras	estas	este	Edfp	/	outras	outro	Iafp
esoutro	ese	ese	Edms	/	outro	outro	Iams
esoutros	eses	ese	Edmp	/	outros	outro	Iamp
esoutra	esa	ese	Edfs	/	outra	outro	Iafs
esoutras	esas	ese	Edfp	/	outras	outro	Iafp
aqueloutro	aquel	aquel	Edms	/	outro	outro	Iams
aqueloutros	aqueles	aquel	Edmp	/	outros	outro	Iamp
aqueloutra	aquela	aquel	Edfs	/	outra	outro	Iafs
aqueloutras	aquelas	aquel	Edfp	/	outras	outro	Iafp

	Constituínte 1				Constituínte 2				Constituínte 3		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
destoutro	de	de	P	/	este	este	Edms	/	outro	outro	Iams
destoutros	de	de	P	/	estes	este	Edmp	/	outros	outro	Iamp
destoutra	de	de	P	/	esta	este	Edfs	/	outra	outro	Iafs
destoutras	de	de	P	/	estas	este	Edfp	/	outras	outro	Iafp
desoutro	de	de	P	/	ese	ese	Edms	/	outro	outro	Iams
desoutros	de	de	P	/	eses	ese	Edmp	/	outros	outro	Iamp
desoutra	de	de	P	/	esa	ese	Edfs	/	outra	outro	Iafs
desoutras	de	de	P	/	esas	ese	Edfp	/	outras	outro	Iafp
daqueloutro	de	de	P	/	aquel	aquel	Edms	/	outro	outro	Iams
daqueloutros	de	de	P	/	aqueles	aquel	Edmp	/	outros	outro	Iamp
daqueloutra	de	de	P	/	aquela	aquel	Edfs	/	outra	outro	Iafs
daqueloutras	de	de	P	/	aquelas	aquel	Edfp	/	outras	outro	Iafp

Se nos fixamos na etiqueta do demostrativo nestas dúas táboas, primeiro ou segundo constituínte dependendo da contracción, constatamos que, fóra os dous últimos caracteres correspondentes aos valores dos atributos de xénero e número, esta é en todos

os casos igual: Ed (*Demostrativo determinante*). Sinalamos esta caracterización morfosintáctica para, no contraste coas etiquetas dos demostrativos masculinos e femininos no lexicón que levan Ea (*Demostrativo determinante/non determinante*), subliñar a intencionalidade presente no establecemento dos constituíntes das distintas contraccións; ou sexa, marcamos en cada un dos compoñentes unicamente a(s) etiqueta(s) que o pode(n) clasificar, sen que a caracterización que propoñemos para ese elemento como formante dunha contracción determinada teña que coincidir forzosamente coa que propoñemos no lexicón para ese mesmo elemento cando aparece illado. Ademais, tampouco ten por que existir coincidencia na etiquetación dun elemento por moito que sexa o mesmo elemento o que apareza en contraccións distintas. Continuando a exemplificación cos demostrativos, na amálgama destes coas preposicións *en* ou *de* é obvio que a súa etiqueta debe coincidir coa que se lle proporciona cando non se integra en ningunha unidade contracta, pois os seus valores son os mesmos xa que, exceptuando o neutro, tanto poden funcionar como *determinantes* coma como *non determinantes*. Así é que, fixémonos na táboa seguinte onde recollemos as amálgamas dos demostrativos con *en*, a caracterización morfosintáctica das formas demostrativas masculinas ou femininas na contracción con preposición difire da dese mesmo demostrativo combinado coas mesmas preposicións cando se unen ao indefinido “outro”, pois neste último caso a mera presenza das formas do indefinido supoñen a aplicación do valor *determinante* ás formas demostrativas:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
neste	en	en	P	/	este	este	Eams
nestes	en	en	P	/	estes	este	Eamp
nesta	en	en	P	/	esta	este	Eafs
nestas	en	en	P	/	estas	este	Eafp
nese	en	en	P	/	ese	ese	Eams
neses	en	en	P	/	eses	ese	Eamp
nesa	en	en	P	/	esa	ese	Eafs
nesas	en	en	P	/	esas	ese	Eafp
naquel	en	en	P	/	aquel	aquel	Eams
naqueles	en	en	P	/	aqueles	aquel	Eamp
naquela	en	en	P	/	aquela	aquel	Eafs
naquelas	en	en	P	/	aquelas	aquel	Eafp
nisto	en	en	P	/	isto	este	Enns
nesto	en	en	P	/	esto	este	Enns
niso	en	en	P	/	iso	ese	Enns
neso	en	en	P	/	eso	ese	Enns
naquilo	en	en	P	/	aquilo	aquel	Enns
naquelo	en	en	P	/	aquelo	aquel	Enns

Naturalmente, as contraccións que non son admitidas pola norma, tal e como observamos con só fixarnos nas últimas filas da anterior táboa, tamén teñen cabida no sistema de XIADA, xa que a súa presenza nos textos do CORGA obriga a implementalas, pois a finalidade fundamental de todos os recursos cuxa elaboración estamos a describir é a etiquetación morfosintáctica deste corpus. Así, non só *nesto*, *neso* e *naquelo* son consideradas no noso sistema contraccións non normativas. A lista abrangue 32 unidades ortográficas. Vexámolas.

No conxunto das unidades contractas non aceptadas pola norma actual podemos distinguir a efectos de descrición varios grupos. O primeiro deles está constituído por seis unidades gráficas que coa publicación da nova edición das *Normas* (RAG/ILG, 2003) deixaron de considerarse aceptables na norma culta, xa que ao deixar de selo o segundo dos compoñentes desas unidades, por extensión, calquera unidade na que ese elemento estea implicado tampouco o é. Falamos, como non, do demostrativo neutro *esto*, *eso* e *aquelo* en combinación coas preposicións *de* e *en*:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
desto	de	de	P	/	esto	este	Enns
deso	de	de	P	/	eso	ese	Enns
daquelo	de	de	P	/	aquelo	aquel	Enns
nesto	en	en	P	/	esto	este	Enns
neso	en	en	P	/	eso	ese	Enns
naquelo	en	en	P	/	aquelo	aquel	Enns

O segundo dos grupos está formado por catro contraccións, as da preposición *pra* co artigo determinado que, a semellanza das do grupo anterior, non se aceptan na norma culta por non considerarse normativa a preposición *pra*¹¹⁰.

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
pró	pra	pra	P	/	o	o	Ddms
prós	pra	pra	P	/	os	o	Ddmp
prá	pra	pra	P	/	a	o	Ddfs
prás	pra	pra	P	/	as	o	Ddfp

¹¹⁰ Nas *Normas* (RAG/ILG, 2003: 182) explicitase que “*normativamente debe optarse por para, pois, aínda que hoxe é minoritaria, é congruente co criterio de darlles preferencia ás formas plenas sobre as reducidas*”; mais no VOLGa (González e Santamarina, 2004: 655, 664) inclúen as entradas *pra*, *prá* e *pró* precedidas do signo * e nos sinónimos, despois do apropiado “*para*” ou “*para a / para o*”, empregan os símbolos =>, co que se está indicando que o sinónimo é preferible sobre o da entrada, pero esta non está condenada e é tolerada nun, en teoría, estándar para o léxico do galego. Na teoría o VOLGa (González e Santamarina, 2004: 5) adaptouse ao novo texto das *Normas* (RAG/ILG, 2003); así que, por quen debemos guiarnos?

O seguinte grupo de contraccións non normativas éo porque estas non se representan nunca na escrita; é o caso das amálgamas dos indefinidos *algo* e *alguén* e mais as do adverbio *algures* coas preposicións *de* e *en*:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
dalgo	de	de	P	/	algo	algo	In00
dalguén	de	de	P	/	alguén	alguén	Inas
dalgures	de	de	P	/	algures	algures	Wn
nalgo	en	en	P	/	algo	algo	In00
nalguén	en	en	P	/	alguén	alguén	Inas
nalgures	en	en	P	/	algures	algures	Wn

Para calquera cuestión relacionada coa elección da etiqueta do indefinido nas contraccións anteriores remitimos á epígrafe correspondente do capítulo 1 sobre o etiquetario de XIADA: 1.3.2.10. Indefinido (I).

A continuación veñen, na nosa relación de unidades contractas non normativas, dúas amálgamas de dous elementos que sempre contraen, a preposición *con* e mais o artigo determinado, só que nesta ocasión a contracción coa forma do artigo feminino resólvese na fala de dúas maneiras distintas, *coa* e *ca*, mentres que na lingua escrita favorécese a forma plena sobre a reducida, porque, ademais (RAG/ILG, 2003: 86), “*desta maneira mantéñense tamén diferentes a contracción de con + a e a conxunción comparativa ca*”¹¹¹:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
ca	con	con	P	/	a	o	Ddfs
cas	con	con	P	/	as	o	Ddfp

Finalmente, as últimas doce unidades contractas non normativas da nosa lista –*cunha, cunhas, cun, cuns, entrambos, entrambas, entrámbolos, entrámbalas e llo, lla, llos, llas*– precisan dunha explicación. *Per se*, non podemos afirmar que se trate de unidades rexeitadas na norma culta, mais dado que no noso sistema a unidade non aparece illada, senón que está ligada indisolublemente a unha forma, un lema e unha etiqueta morfosintáctica, si estamos en disposición de realizar esa aseveración.

¹¹¹ É verdade que así a diferenza entre unha e outra é clara, mais en galego existen outros casos de ambigüidades segmentais ben comúns que crean aínda máis problemas ca os desta contracción, pensemos por exemplo en *polos*.

No VOLGa aparece esta contracción de preposición e artigo femino como variante non recomendada, mais si tolerada. En XIADA, como vemos, indícase que é unha forma non normativa porque tomamos como referencia as *Normas* (RAG/ILG, 2003: 85-86).

Comentabamos máis arriba, en relación coas posibles etiquetas do demostrativo nas distintas contraccións en que actúa, que cada un dos elementos dunha amálgama vai ser caracterizado morfoloxicamente coas etiquetas que lle correspondan atendendo precisamente á súa integración na amálgama. Pois ben, no caso da preposición *con* contraída con *un, uns, unha, unhas* non nos enfrontamos con catro unidades gráficas senón con doce. Catro delas, normativas, concirnen á contracción co artigo indeterminado; outras catro, non normativas, proceden da contracción co indefinido non determinante¹¹² e, finalmente, das catro restantes, as correspondentes á súa análise como contraccións co numeral, dúas son consideradas normativas –cando o numeral se etiqueta como *determinante*– e outras dúas, non normativas –cando o numeral se etiqueta como *non determinante*–. Vexamos entón a caracterización concreta destas doce unidades, centrándonos na etiquetación do segundo compoñente, e discriminando entre unidades normativas e unidades non normativas¹¹³:

Contracción	Normatividade da contracción	Constituínte 1				Constituínte 2		
		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
cunha	non	con	con	P	/	unha	un	Infs
cunha	non	con	con	P	/	unha	un	Ncnfs
cunha	si	con	con	P	/	unha	un	Difs
cunha	si	con	con	P	/	unha	un	Ncdfs
cunhas	non	con	con	P	/	unhas	un	Infp
cunhas	si	con	con	P	/	unhas	un	Difp
cun	non	con	con	P	/	un	un	Inms
cun	non	con	con	P	/	un	un	Ncnms
cun	si	con	con	P	/	un	un	Dims
cun	si	con	con	P	/	un	un	Ncdms
cuns	non	con	con	P	/	uns	un	Inmp
cuns	si	con	con	P	/	uns	un	Dimp

Caso moi semellante a este que acabamos de ver constitúeo a contracción do pronome de terceira dativo *lle* cos de acusativo, tamén de terceira, cando o dativo non

¹¹² No capítulo seguinte abordaremos, entre outras cuestións, o criterio establecido en XIADA para diferenciar nas formas *un, uns, unha, unhas* a súa adscripción ás categorías Artigo, Indefinido e Numeral.

¹¹³ Escollemos representar na táboa nunha columna a normatividade das formas contractas porque nos parecía clara e ademais é a que figura no arquivo pertencente ás contraccións do léxico descargable de XIADA (Rojo et al., 2009), non obstante, como explicaremos máis adiante, computacionalmente realízase doutra maneira.

ten valor singular senón plural. Sabemos que no galego estándar debe observarse a distinción no dativo de terceira persoa entre unha forma singular *lle* e outra plural *lles*. Con todo, nos textos actuais aínda, en ocasións, se reflicte o descoñecemento da forma plural *lles* (Álvarez, Regueira e Monteagudo, 1986: 164) ao constatar a utilización de *lle* para singular e para plural; en consecuencia, a forma *lle*, maioritariamente singular, é etiquetada en casos coma os seguintes co valor plural:

Habitualmente están dispostos a participar de xeito desinteresado en proxectos que lle gusten e sexan serios. [CORGAetq: ANT2005-03-03/72]

Este centro, ademais de achegarlle formación ás persoas con discapacidade, terá tamén unha área de formación para profesionais e os medios necesarios para a utilización das novas tecnoloxías. [CORGAetq: GH2007-03-17/32]

Cando o pronome *lle* contrae co de acusativo de terceira realízase sempre como *llo*, *lla*, *llos*, *llas*, mais o *lle* da contracción, dá igual que apareza illado ou enclítico a unha forma verbal, tanto pode corresponder á etiqueta Rad3as (*Pronome, átono, dativo, masculino/feminino, singular*) como á de Rad3ap (*Pronome, átono, dativo, masculino/feminino, plural*). Así, *lle* nos seguintes exemplos, posúe valor singular nas dúas primeiras secuencias, mentres que pola contra nas tres últimas actúa como plural:

Que llas saquen á dirección de costas e *llelas* devolvan ós paisanos, os seus lexítimos propietarios. [CORGA: CG2000-11-05/23]

O deportivista Sergio avoga por afrontar o partido de mañá contra o Barcelona coa intención de desfrutar porque a situación *llo* permite [CORGAetq: GH2007-03-30/73]

Tamén pasa cando un de nós vai ver unha película de medo e despois lla conta ós outros. [CORGA: X. Miranda, *A neve e a cadeira*, Nigra, 1994]

Eses e outros contos aprendéraos máis tarde, cando marchara á vila a servir, e llos contaban ás cativas da familia para dormiren. [CORGA: M^a. X. Queizán, *O solpor da cupletista*, Nigra, 1995]

Rexedor: De tanto llos chiscar ás mulleres doutros. [CORGA: R. Vidal Bolaño, *A burla do galo ou Vida e amores de Don Esmeraldino da Cámara Mello de Lima Vizconde de Ribeirinha e Galo de Portugal*, IGAEM, 2000]

Consecuentemente, no noso sistema establécese tamén esa diferenza na caracterización das unidades nas que o *lle* está contraído ao considerar que este primeiro compoñente presenta dúas análises posibles: nunha delas, na máis frecuente e na que dá lugar ás amálgamas normativas, caracterízase como Rad3as mentres que na outra, na

máis esporádica e cuxa presenza implica que as contraccións resultantes son non normativas, márcase coa etiqueta Rad3ap. Vexamos as súas características:

Contracción	Normatividade da contracción	Constituínte 1				Constituínte 2		
		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
llo	si	lle	lle	Rad3as	/	o	o	Raa3ms
llo	non	lle	lle	Rad3ap	/	o	o	Raa3ms
llos	si	lle	lle	Rad3as	/	os	o	Raa3mp
llos	non	lle	lle	Rad3ap	/	os	o	Raa3mp
lla	si	lle	lle	Rad3as	/	a	o	Raa3fs
lla	non	lle	lle	Rad3ap	/	a	o	Raa3fs
llas	si	lle	lle	Rad3as	/	as	o	Raa3fp
llas	non	lle	lle	Rad3ap	/	as	o	Raa3fp

Por último, para rematar coa descrición das contraccións non normativas, e tal e como comentamos en relación co número de constituíntes que pode posuír unha contracción en galego, en XIADA consideramos que as unidades ortográficas *entrambos*, *entrambas*, *entrámbolos* e *entrámbalas* presentan ambigüidade segmental. Que significa isto? Pois ben, con esta expresión pretendemos indicar que son *tokens* que non presentan unha única alternativa de análise senón varias, entendendo por alternativa cada unha das posibles segmentacións da unidade ortográfica. Pensemos por exemplo en *cóntanos*¹¹⁴. Que elementos lingüísticos se agachan baixo esa unidade gráfica? Evidentemente, unha forma verbal e un pronome enclítico; pero, cales? É ambiguo. Nunha análise podemos ver a terceira persoa singular do presente de indicativo ou incluso a segunda singular do imperativo *conta* seguida do pronome de acusativo ou dativo *nos*, en tanto que noutra análise estamos ante a terceira persoa do plural do presente de indicativo *contan* e mais o pronome átono de acusativo masculino plural *os*:

cóntanos:

- 1) conta + nos
- 2) contan + os

¹¹⁴ Non temos por que recorrer ao manido *polos* para exemplificar, pois en galego danse bastantes casos de ambigüidades segmentais; pensemos, senón, no que ocorre coas formas verbais regulares de terceira persoa singular cando levan enclítico o pronome *nos* ou nas de terceira plural cando levan enclítico o pronome de acusativo masculino plural *os*: *cántanos* (*canta* + *nos* ou *cantan* + *os*), *cantábanos*, *cantáranos*, *cantarianos*, *cántenos*, *cantásenos*...

En definitiva, *entrambos*, *entrambas*, *entrámbolos* e *entrámbalas* son unidades gráficas nas que a identificación das correspondentes unidades gramaticais é ambigua debido a que se poden segmentar de máis dunha maneira. Así, nunha das análises, á unidade gráfica *entrambos*, *entrambas* –categorialmente indefinido– correspóndelle unha única unidade gramatical, feito polo que está introducida no lexicón, e o emprego da segunda forma do artigo coas formas dese indefinido dá lugar ás contraccións normativas *entrámbolos*, *entrámbalas* coa seguinte análise:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
entrámbolos	entrambos	entrambos	Idmp	/	los	o	Ddmp
entrámbalas	entrambas	entrambos	Idfp	/	las	o	Ddfp

Na segunda das análises posibles, as unidades ortográficas *entrambos*, *entrambas* correspóndense coas contraccións non normativas da preposición *entre* e mais o indefinido *ambos*, sen cabida no lexicón e engrosando, polo tanto, a relación de contraccións, e dentro destas, a lista das non correctas en galego hoxe por hoxe; o mesmo vale para a amálgama destas coas segundas formas do artigo:

	Constituínte 1				Constituínte 2				Constituínte 3		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
entrambos	entre	entre	P	/	ambos	ambos	Idmp				
entrambas	entre	entre	P	/	ambas	ambos	Idfp				
entrámbolos	entre	entre	P	/	ambos	ambos	Idmp	/	los	o	Ddmp
entrámbalas	entre	entre	P	/	ambas	ambos	Idfp	/	las	o	Ddfp

O inventario de contraccións de XIADA engrósano tamén unidades non consideradas en xeral neste apartado nos estudos gramaticais. Referímonos, por suposto, ás aglutinacións nas que está envolta a segunda forma do artigo cando, exceptuando as formas verbais, aparece en combinación coas preposicións *por* e *tras*, os indefinidos *ambos*, *entrambos* e *todos*, a conxunción *mais*, os pronomes tónicos *nós* e *vós* e o adverbio *u*.

De todos é coñecido que a variante alomórfica *lo*, *la*, *los*, *las* é de uso obrigatorio (RAG/ILG, 2003: 83-85) tras da preposición *por* e do adverbio *u*, así como que tamén se pode representar na escrita despois de infinitivo, das formas verbais conxugadas que rematan en *-r* ou *-s*, dos pronomes enclíticos *nos*, *vos* e *lles*, dos pronomes tónicos *nós* e *vós* cando van seguidos dunha precisión numérica, despois de *ambos*, *entrambos* e *todos* cos seus femininos, da preposición *tras* e da conxunción copulativa (*e*) *mais*.

Se comparamos os elementos que se combinan ou se poden combinar coa segunda forma do artigo descritos nos dous últimos parágrafos, comprobamos que

existe unha discrepancia entre eles: a norma establece a posibilidade de utilizar o alomorfo *lo, la, los, las* tras infinitivo, formas verbais conxugadas rematadas en *-r* ou *-s* e cos pronomes enclíticos *nos, vos* e *lles*, mentres que nós non incluímos nas nosas contraccións estas tres opcións. A explicación relativa á ausencia das formas verbais con segunda forma do artigo establecémosla ao comezo deste apartado por considerar que o sistema empregado para implementar tales aglutinacións debía seguir outro proceso, o cal veremos máis adiante. Polo que respecta á combinación do artigo cos pronomes átonos *nos, vos* e *lles*, esta dáse sempre enclítica a unha forma verbal, motivo polo que será tratada tamén nesa sección.

Que diferenza existe entón no noso sistema entre as contraccións do artigo coas preposicións *de, en* ou *a* e as que teñen lugar con *por* e *tras*? É na caracterización do segundo elemento onde se observa a discrepancia, pois mentres para as primeiras a forma do segundo constituínte é *o, a, os, as* nestas últimas considérase que a forma é *lo, la, los, las*. Vexamos por exemplo a etiquetación empregada nas amálgamas con *por*:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
polo	por	por	P	/	lo	o	Ddms
pola	por	por	P	/	la	o	Ddfs
polos	por	por	P	/	los	o	Ddmp
polas	por	por	P	/	las	o	Ddfp

A mesma caracterización ten lugar nas contraccións cos indefinidos *ambos, entrambos* e *todos* para o artigo:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
ámbolos	ambos	ambos	Idmp	/	los	o	Ddmp
ámbalas	ambas	ambos	Idfp	/	las	o	Ddfp
entrámbolos	entrambos	entrambos	Idmp	/	los	o	Ddmp
entrámbalas	entrambas	entrambos	Idfp	/	las	o	Ddfp
tódolos	todos	todo	Idmp	/	los	o	Ddmp
tódalas	todas	todo	Idfp	/	las	o	Ddfp

O emprego ou non de trazo na representación gráfica da segunda forma do artigo non supón computacionalmente ningunha diferenza; as unidades contractas sen trazo incorpóranse sen el e as que o utilizan recóllense con el, pero sen que este factor implique ningunha marca a maiores nin inflúa na caracterización. Contrástese, por exemplo, a marcade do segundo constituínte nas contraccións coa conxunción coas do adverbio ou as dos pronomes tónicos:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
mailo	mais	mais	Cc	/	lo	o	Ddms
maila	mais	mais	Cc	/	la	o	Ddfs
mailos	mais	mais	Cc	/	los	o	Ddmp
mailas	mais	mais	Cc	/	las	o	Ddfp
u-lo	u ¹¹⁵	u	Wn	/	lo	o	Ddms
u-la	u	u	Wn	/	la	o	Ddfs
u-los	u	u	Wn	/	los	o	Ddmp
u-las	u	u	Wn	/	las	o	Ddfp
nó-los	nós	nós	Rtr1mp	/	los	o	Ddmp
nó-las	nós	nós	Rtr1fp	/	las	o	Ddfp
vó-los	vós	vós	Rtr2mp	/	los	o	Ddmp
vó-las	vós	vós	Rtr2fp	/	las	o	Ddfp

A mesma distinción na *forma* entre as denominadas primeira e segunda forma levámola á práctica nas contraccións do pronome de acusativo cos de dativo. Así, mentres a forma do segundo constituínte nas contraccións con *me*, *che* e *lle* é *o*, *a*, *os*, *as*, nas amálgamas con *nos*, *vos* e *lles* é *lo*, *la*, *los*, *las*:

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
mo	me	me	Rad1as	/	o	o	Raa3ms
mos	me	me	Rad1as	/	os	o	Raa3mp
ma	me	me	Rad1as	/	a	o	Raa3fs
mas	me	me	Rad1as	/	as	o	Raa3fp
cho	che	che	Rad2as	/	o	o	Raa3ms
chos	che	che	Rad2as	/	os	o	Raa3mp
cha	che	che	Rad2as	/	a	o	Raa3fs
chas	che	che	Rad2as	/	as	o	Raa3fp
llo	lle	lle	Rad3as	/	o	o	Raa3ms
llos	lle	lle	Rad3as	/	os	o	Raa3mp
lla	lle	lle	Rad3as	/	a	o	Raa3fs
llas	lle	lle	Rad3as	/	as	o	Raa3fp
llo	lle	lle	Rad3ap	/	o	o	Raa3ms
llos	lle	lle	Rad3ap	/	os	o	Raa3mp
lla	lle	lle	Rad3ap	/	a	o	Raa3fs
llas	lle	lle	Rad3ap	/	as	o	Raa3fp
nolo	nos	nos	Rad1ap	/	lo	o	Raa3ms
nolos	nos	nos	Rad1ap	/	los	o	Raa3mp
nola	nos	nos	Rad1ap	/	la	o	Raa3fs
nolas	nos	nos	Rad1ap	/	las	o	Raa3fp
volos	vos	vos	Rad2ap	/	lo	o	Raa3ms
volos	vos	vos	Rad2ap	/	los	o	Raa3mp

¹¹⁵ Supón unha incoherencia no noso sistema que a palabra á que se une a segunda forma do artigo ou o alomorfo do pronome de acusativo recupere sempre graficamente na *forma* o carácter asimilado (-r ou -s), mentres que no caso do adverbio *u* (<*ub) non ocorre así. As distintas obras lexicográficas –DRAG (1997), Ir Indo (2004), Xerais (2000, 2009) e VOLGa (2004)– achegan como entrada para este adverbio *u*, co que debemos entender que esa é tanto a forma como o lema do primeiro compoñente. Con todo, a cuestión intríganos, pois se este adverbio só pode utilizarse seguido das segundas formas do artigo ou do pronome, por que posúe unha entrada propia e illada? E por que, se a súa expresión actual é *u*, insistir en que hai que empregalo só coas segundas formas?

	Constituínte 1				Constituínte 2		
	<i>forma</i>	<i>lema</i>	<i>etiqueta</i>		<i>forma</i>	<i>lema</i>	<i>etiqueta</i>
vola	vos	vos	Rad2ap	/	la	o	Raa3fs
volas	vos	vos	Rad2ap	/	las	o	Raa3fp
llelo	lles	lle	Rad3ap	/	lo	o	Raa3ms
llelos	lles	lle	Rad3ap	/	los	o	Raa3mp
llela	lles	lle	Rad3ap	/	la	o	Raa3fs
llelas	lles	lle	Rad3ap	/	las	o	Raa3fp

Acabamos de ver como se tratan dende o punto de vista lingüístico as contraccións en XIADA, mais fáltanos describir como se organiza computacionalmente toda esa información, aspecto que trataremos a continuación, aínda que dun xeito breve.

Nunha base de datos, externa ao lexicón pero accesible ao preprocesador e etiquetador, figuran varias táboas. Nunha delas rexístranse todas as secuencias das contraccións cun identificador e a indicación de se a secuencia é normativa ou non; nesta táboa aparecen todas as contraccións que integramos en XIADA, en total 203, mais nela non recollemos unha secuencia por cada posibilidade de etiquetación, senón só a unidade gráfica (*ó, á, ós, ás, do, da, dos, das*, etc.). Por exemplo, aparece unha soa vez a secuencia *cunha*, para a cal vimos máis arriba que existen catro posibilidades de análise.

Noutra táboa independente introducimos os constituíntes de cada contracción e para cada compoñente incluimos os datos relativos á forma, ás etiquetas posibles, ao lema e á normatividade do compoñente. Respecto a isto último, lembremos que unha determinada contracción *per se* pode ser normativa (*ó, ós, á, ás, ao, aos...*) ou non normativa (*desto, nesto, ca, dalguén...*), mais que tamén aquí existe a ambigüidade, pois unha mesma secuencia contracta, dependendo da súa análise, pode alternar nestes valores. Vímolos coa contracción *cunha*, a cal en principio non podemos caracterizar como non normativa na táboa xeral de secuencias, posto que presenta análises que si o son. Como proceder entón nestas ocasións? Pois marcando como non normativo o compoñente que provoca que a contracción non se adecúe á norma actual; neste caso a pesar de que *cunha* é normativa, o compoñente que ocupa a segunda posición nesa secuencia cando se asocia coas etiquetas Infs (*Indefinido, non determinante, feminino, singular*) e Ncnfs (*Numeral, cardinal, non determinante, feminino, singular*) di que a secuencia non é correcta normativamente.

Finalmente, con outra táboa conectamos a información das secuencias coa dos compoñentes, de xeito que se relaciona o identificador da contracción cos identificadores dos compoñentes, coa posición que ocupa cada un dos compoñentes nas distintas secuencias e co número de análises posibles para cada compoñente. Isto permítenos que, por exemplo, botemos man do mesmo identificador para relacionar todas as contraccións nas que está implicada a preposición *de*, poñamos por caso, xa que esta presenta sempre a mesma forma, etiqueta, lema, normatividade e número de análises.

Agora xa sabemos en que lugar e de que maneira se caracterizan as contraccións no noso sistema, pero para completar a descrición fáltanos aínda ver en que momento do proceso se singulariza a unidade contraída e de que xeito se representa esa análise no documento etiquetado; non obstante, preferimos postergar estes dous últimos aspectos para cando xa teñamos explicado o tratamento dos enclíticos, pois son comúns tanto ao fenómeno das amálgamas como ao da énclise.

2.4. Un caso especial de unidades gramaticais complexas: formas verbais con enclíticos

2.4.1. EAGLES e sistemas do SLI-TECTRA e SLI-CTAG

EAGLES (Monachini e Calzolari, 1996) propón que no nivel L2b, aquel no que se integran as peculiaridades lingüísticas exclusivas dunhas poucas linguas, figure un atributo *clíticos* con cuxos valores booleanos *clíticos*, *non clíticos* se codifique a presenza de pronomes enclíticos nunha forma verbal. Na práctica, no nivel morfolóxico a aplicación de EUROTRA para o castelán non recolle este atributo e na versión portuguesa de GENELEX tampouco o aplicaron.

É na aplicación italiana, recollida tamén na proposta de EAGLES, onde atopamos o xerme do tratamento que anos máis tarde se desenvolverá para o galego, con cambios importantes entre eles, nos sistemas do SLI-TECTRA, SLI-CTAG e XIADA:

In Italian, pronouns (pronominal particles) can accompany verbs in order to form the:

1. Pronominal form of the verb;
2. Reflexive form; and the

3. Reciprocal form.

As far as the lexicon is concerned, the possibility of taking clitics of different types (reflexive, reciprocal, etc.) has to be encoded in the entry, but depends on the syntactic type of the verb; these problems should therefore be dealt with on the syntactic level.

In these forms, the unstressed variant of the pronoun is, in general, separated graphically by the verb (it precedes the verb), except with the infinitive, the gerund and the imperative form and sometimes (rarely) with the past participle:

'egli si lava'

'lavarsi', 'lavandosi', 'lavati', 'lavatosi'.

At the corpus level, a clitic pronoun can be attached to the verb in three situations, when it expresses:

1. The direct object: 'dirlo' ('dire' + 'lo') -- 'to say it';
2. The indirect object: 'dirgli' ('dire' + 'gli') -- 'to say to him'; or
3. A place adverb: 'andarci' ('andare' + 'ci') -- 'to go there'.

Sequences of a verb with more than one pronoun in a unique graphical form are found if both the direct object and the indirect object of a pronominal verb form are represented by a clitic pronoun, e.g. 'dandomelo' ('dando' + 'me' + 'lo') -- 'giving it to me'. These types of compounds can present the phenomenon of epenthesis, i.e. the insertion of a letter for euphonic reasons, e.g. 'dandoglielo' ('dando' + 'gli' + 'e'(epenthetic) + 'lo') -- 'giving it to him'.

As to the dictionary encoding, 'verb-pronoun' compounds have to be represented; the problem of encoding specific 'verb-more-than-one-pronoun' compounds only concerns corpus encoding.

The strategy adopted in tagging these is to assign different tags to the different parts forming the word-token, with a special mark which maintains the graphical links, thus permitting the recovery of the unique graphical form. [Monachini e Calzolari, 1996]

Tras ler esta cita varios son os puntos sobre os que nos gustaría incidir:

1. En galego prodúcese a énclese non só co infinitivo, xerundio e imperativo senón con calquera forma verbal, exceptuando o participio. Ademais, pode empregarse a segunda forma do artigo con calquera forma verbal rematada en *-r* ou *-s*, novamente menos co participio, e a efectos de codificación este fenómeno debe tratarse igual ca énclese.

2. O problema para codificar o composto “forma verbal + pronome enclítico” é exactamente idéntico ao que existe para codificar os compostos “forma verbal + dous/tres... pronomes enclíticos”. Na nosa opinión ambos deben resolverse no dominio

do corpus e non no do lexicón, xa que non existe unha categoría gramatical “verbo-pronome” na que poidamos integrar estes falsos compostos.

3. O asignar diferentes etiquetas aos constituíntes dun composto “verbo + pronome enclítico” é adecuado, así como tamén o é permitir a recuperación da unidade gráfica, mais o feito de introducir unha marca específica para vincular os constituíntes coa unidade gráfica xa non o é tanto se provoca a multiplicación de etiquetas.

Sabemos que non é tarefa fácil tratar con unidades gráficas que acollen máis dunha unidade lingüística, e a dificultade incrementase aínda máis se, como ocorre co fenómeno da éncise ou co emprego da segunda forma do artigo cos verbos, un dos constituíntes presenta unha variación tal na súa forma, etiqueta e lema que case podemos falar dun número ilimitado de elementos que poden constituír ese compoñente. Con todo, non se trata dunha tarefa inabarcable ou imposible, só é preciso afrontala dende unha perspectiva diferente ás que manexamos ata agora.

Como explicamos en Barcala et al. (2007) existen dúas maneiras de tentar resolver os casos nos que un mesmo *token* contén información sobre dous ou máis compoñentes lingüísticos:

1^a) Tratar os *tokens* no seu conxunto e estender o etiquetario para representar estes fenómenos. Por exemplo, *colléullelo* podería ser etiquetado como Vei30s+Rad3ap+Raa3ms (*Verbo, pretérito, indicativo, 3^a, xénero non aplica, singular + Pronome, átono, dativo, 3^a, masculino/feminino, plural + Pronome, átono, acusativo, 3^a, masculino, singular*). Outra alternativa consistiría en etiquetar *colléullelo* como Vei30s2, onde o último 2 indica que a forma verbal leva dous pronomes enclíticos.

2^a) Segmentar os *tokens* compostos, separando os seus compoñentes. Por exemplo, a mesma palabra *colléullelo* pode ser separada en tres partes: *colléu, lle* e *lo*. Nestes casos os compoñentes poden ser etiquetados separadamente como Vei30s, Rad3ap e Raa3ms, respectivamente.

Aínda que EAGLES apunta á utilización deste segundo enfoque cando o fenómeno é recorrente nunha lingua, moitos sistemas están baseados no primeiro enfoque (Aguirre Moreno et al., 2001, 2002b; Carreras et al., 2004). É o máis sinxelo, pois non é necesario modificar os etiquetadores existentes e actúa ben en linguas que apenas presentan casos deste tipo, coma o inglés ou o francés, pero presenta serios

problemas con aquelas linguas que, coma a nosa, teñen numerosas ocorrencias de varios elementos lingüísticos na mesma palabra. Nestes casos, o tamaño do etiquetario increméntase considerablemente e ás veces as tarefas de etiquetación son pouco prácticas debido a que necesitan un corpus de adestramento extremadamente amplo para obter un comportamento aceptable do etiquetador. En xeral, tampouco os traballos que seguen esta opción explican como solucionar outros problemas relacionados coa etiquetación: como asignar lemas aos *tokens* compostos, como empregar o sistema dentro doutro máis complexo, por exemplo cun tradutor automático ou cun analizador sintáctico, etc., cando estes problemas son triviais se se emprega a segunda alternativa.

As taras que acabamos de sinalar podemos atribuírllelas aos sistemas do SLI-TECTRA e do SLI-CTAG, pois se recuperamos só un exemplo de unidade gráfica amalgamada vemos que o primeiro considera a existencia de etiquetas compostas e incorre na incongruencia de proporcionar dous lemas para unha única unidade gráfica (Aguirre Moreno et al., 2002b) :

```
<mor lema="contar" lema2="se" cat="VIPRS3_PPS3AR"/> <ort>cóntase</ort> </ling>
```

mentres que no SLI-CTAG, a pesar de non empregar etiquetas compostas, non se diferencian as formas correspondentes a cada un dos elementos gramaticais nin se diferencia nas análises dos elementos, polo menos no formato escollido (Gómez Guinovart e López Fernández, 2009), cando estes están illados de cando están amalgamados, posto que o signo ^ marca sempre o inicio de cada análise, sexa gráfica ou sexa lingüística, é dicir, exista ou non correspondencia entre unidade gráfica e unidade lingüística. Ademais, coma no exemplo que segue, a lectura que facemos ao extraer a información é que á forma ortográfica amalgamada, *lévanse* neste caso, se lle asigna o lema da primeira unidade lingüística, *levar*, e ese dato é falso xa que é á forma *levan* á que lle corresponde o lema *levar*, mais non a *lévanse*:

```
^lévanse/levar_VIC3P00      ^~/se_P3CNR0
```

2.4.2. Tratamento no sistema de XIADA

2.4.2.1. Xeneralidades

Xa sabemos, porque así o vimos coas contraccións, que o sistema desenvolvido en XIADA segue o segundo enfoque, mais o camiño que percorremos é distinto no caso

das formas verbais seguidas de pronomes enclíticos, a pesar de que o resultado final é o mesmo: *token* composto desagregado nos seus compoñentes pertinentes, facendo constar para cada constituínte a forma, a etiqueta e o lema, sen perder de vista a unidade ortográfica. Por exemplo, a saída do etiquetador de XIADA, nun formato de texto semellante aos empregados nos sistemas do SLI, para unha unidade coma *Produciraas*, extraída da oración “*_Produciraas se a sociedade llo permite.*” da versión 2.4 do CORGAetq é a seguinte:

```
<análise_unidade>
  <unidade>Produciraas</unidade>
  <constituínte>
    <forma>producirá</forma>
    <etiqueta>Vfi30s</etiqueta>
    <lema>producir</lema>
  </constituínte>
  <constituínte>
    <forma>as</forma>
    <etiqueta>Raa3fp</etiqueta>
    <lema>o</lema>
  </constituínte>
</análise_unidade>
```

Para representar a etiquetación optamos, despois de estudar pormenorizadamente os estándares CES (Ide, 2010), XCES (Vassar College e Loria, 2008) e TEI (Burnard e Bauman, 2009), por definir un XML propio¹¹⁶ e, tal como se observa, empregamos o termo *análise_unidade* para acoller a caracterización morfolóxica da unidade gráfica, sexa esta simple, multipalabra ou composta por varios elementos gramaticais. Logo identificamos a unidade que imos analizar, co cal sempre poderemos acceder dun xeito cómodo a esta información, para a continuación desagregar a unidade en tantos compoñentes como posúa e analizar cada un deles en *constituínte*, que á súa vez contará cunha *forma*, cunha *etiqueta* e cun *lema*; cando remata a análise do primeiro constituínte séguese inmediatamente a do segundo, a do terceiro, etc., ata rematar a análise completa do *token* composto. Se prestamos atención aos detalles, podemos destacar varias particularidades desta primeira exemplificación: primeira, a unidade aparece en inicio de oración, co cal o seu primeiro carácter é maiúsculo, pero na *forma* do primeiro *constituínte* aparece xa con minúscula; en segundo lugar, a *forma* “producirá” do

¹¹⁶ Deste xeito empregamos só as etiquetas que realmente necesitamos, e ademais, ao usalas en galego, familiarizámonos antes coa linguaxe de codificación e promocionamos a nosa lingua.

primeiro *constituínte* aparece con acento cando na unidade non o ten e, finalmente, a fácil identificación e etiquetación dos dous constituíntes, pois non presentan ambigüidade de ningún tipo.

Non obstante, a caracterización dunha forma verbal con pronomes enclíticos pode complicarse moito máis se entra en xogo a ambigüidade. É dicir, a análise morfolóxica dunha unidade gráfica composta por máis dun elemento gramatical pode non só ofrecer máis dunha posibilidade á hora de asignarlle a etiqueta a algún dos seus compoñentes, senón que tamén pode ofrecer a posibilidade, en ocasións, de presentar alternativas de análise distintas nas que os elementos desagregados varían totalmente dunha a outra. Trátase dunha situación bastante máis habitual en galego do que nun principio puideramos pensar, e é o que ocorre, por exemplo, coa unidade ortográfica *obríganos*, para a cal a saída do texto preprocesado é a seguinte¹¹⁷:

```

<análise_unidade>
<unidade>obríganos</unidade>
<alternativas>
<alternativa>
<constituínte>
<forma>obriga</forma>
<etq_ lema>
<etiqueta>V0m20s</etiqueta>
<lema>obrigar</lema>
</etq_ lema>
<etq_ lema>
<etiqueta>Vpi30s</etiqueta>
<lema>obrigar</lema>
</etq_ lema>
</constituínte>
<constituínte>
<forma>nos</forma>
<etq_ lema>
<etiqueta>Raa1ap</etiqueta>
<lema>nos</lema>
</etq_ lema>
<etq_ lema>
<etiqueta>Raa1fp</etiqueta>
<lema>nos</lema>
</etq_ lema>
<etq_ lema>
<etiqueta>Raa1mp</etiqueta>
<lema>nos</lema>
</etq_ lema>
<etq_ lema>
<etiqueta>Rad1ap</etiqueta>

```

¹¹⁷ Destaco co formato en letra grosa as dúas alternativas de análise posibles para esta unidade, entendendo que a alternativa implica non que os constituíntes presenten varias etiquetas, senón que existe máis dunha segmentación en constituíntes distinta.

```

    <lema>nos</lema>
  </etq_lema>
  <etq_lema>
    <etiqueta>Rad1fp</etiqueta>
    <lema>nos</lema>
  </etq_lema>
  <etq_lema>
    <etiqueta>Rad1mp</etiqueta>
    <lema>nos</lema>
  </etq_lema>
  <etq_lema>
    <etiqueta>Raf1ap</etiqueta>
    <lema>nos</lema>
  </etq_lema>
  <etq_lema>
    <etiqueta>Raf1fp</etiqueta>
    <lema>nos</lema>
  </etq_lema>
  <etq_lema>
    <etiqueta>Raf1mp</etiqueta>
    <lema>nos</lema>
  </etq_lema>
</constituínte>
</alternativa>
<alternativa>
  <constituínte>
    <forma>obrigan</forma>
    <etq_lema>
      <etiqueta>Vpi30p</etiqueta>
      <lema>obrigar</lema>
    </etq_lema>
  </constituínte>
  <constituínte>
    <forma>os</forma>
    <etq_lema>
      <etiqueta>Raa3mp</etiqueta>
      <lema>o</lema>
    </etq_lema>
  </constituínte>
</alternativa>
</alternativas>
</análise_unidade>

```

Nesta saída comprobamos xa en primeira instancia a existencia de dúas alternativas de análise posibles: *obriga* + *nos* fronte a *obrigan* + *os*. Mentres que a segunda alternativa, a da terceira persoa do presente de indicativo do lema *obrigar* (*obrigan*) co pronome enclítico de terceira acusativo masculino plural (*os*), non presenta duplicidades na análise, pois os seus constituíntes só posúen unha etiqueta, na primeira alternativa cada un dos constituíntes presenta máis dunha posibilidade de análise. Así, a forma *obriga* pode caracterizarse como imperativo segunda singular ou como terceira singular do presente de indicativo, cuxo lema en ambos os casos é *obrigar*. Pola súa banda, para o segundo constituínte, o pronome de primeira plural, lema *nos*, tamén se

achegan todas as posibilidades de análise: a súa adscrición ao acusativo, dativo ou formante léxico coas variacións pertinentes no xénero para o masculino, feminino ou indeterminado; en total 9 etiquetas factibles.

Unha vez que se produce o posprocesado do texto que queremos etiquetar, o noso etiquetador, principalmente estatístico, escolle a alternativa que lle parece máis probable, pero para chegar aí tivemos antes que realizar unha serie de tarefas que lle permiten proporcionarnos eses resultados. Vexamos cales son.

2.4.2.2. Lexicóns verbais: sublexicón *principal* e sublexicón *clíticos*

En primeiro lugar é necesario que o sistema conte cun lexicón de formas verbais que poidan levar enclíticos, mais non confundamos o poder levalos con levalos. Como vimos con anterioridade neste mesmo capítulo, implementamos unha dobre conxugación para cada un dos verbos que integran o lexicón:

1.a) Conxugación verbal no sublexicón *principal*

Por unha banda, elaboramos a conxugación do paradigma verbal, semellante á que pode aparecer en calquera manual, só que a integramos no noso sistema mediante grupos formais. Os datos referentes a esta conxugación gárdanse no sublexicón *principal* e achegan información sobre a raíz verbal, o lema e o grupo ao que remitimos esa raíz, sabendo que neste se lle van proporcionar as desinencias e as etiquetas para cada unha das formas conxugadas. Por exemplo, introducimos no lexicón *principal* o verbo *contar* cos seguintes detalles:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	cont	V	V1	contar	V	si	volga_2004

Logo, no grupo V1 figuran as desinencias e etiquetas correspondentes a cada unha das formas verbais resultantes da conxugación de *contar*; por exemplo, as correspondentes ao presente de indicativo permítennos caracterizar e xerar as unidades gráficas *conto*, *contas*, *conta*, *contamos*, *contades*, *contan*:

Desinencia	Subetiqueta	Seguinte grupo	Normativa
o	pi10s		si
as	pi20s		si
a	pi30s		si
amos	pi10p		si
ades	pi20p		si
an	pi30p		si

1.b) Conxugación verbal con enclíticos no sublexicón *clíticos*

Por outra banda, creamos unha conxugación paralela á primeira que se garda no sublexicón *clíticos* e que está formada polo paradigma conxugado completo de cada verbo presente no lexicón, pero coa particularidade de que as formas verbais aí integradas son as que poden ir acompañadas dun ou máis enclíticos, incluíndo nestes a segunda forma do artigo. Así, para a introdución no lexicón *clíticos* do verbo *contar* damos os seguintes datos:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	clíticos	cont	V	Vc1a	contar	V	si	volga 2004
xeral	clíticos	cónt	V	Vc1at	contar	V	si	volga 2004

Logo, no grupo Vc1a achegamos as terminacións verbais coas correspondentes etiquetas para que o sistema poida recoñecer e xerar as formas verbais arrizotónicas cando se presentan en combinación con enclíticos ou segunda forma do artigo, en tanto que no Vc1at se achega a mesma información para as formas verbais rizotónicas cando se atopan en combinación con pronomes enclíticos ou segunda forma do artigo. Deste xeito, por exemplo, para o recoñecemento ou xeración das formas verbais do presente de indicativo cando se amalgaman con clíticos, introducimos as seguintes terminacións:

[Raíz]	Grupo	Desinencia	Subetiqueta	Normativa
[CONT]	Vc1a	ámos	pi10p	si
		ámo	pi10p	si
		ádes	pi20p	si
		áde	pi20p	si
[CÓNT]	Vc1at	o	pi10s	si
		as	pi20s	si
		a	pi20s	si
		a	pi30s	si
		an	pi30p	si

Así, sempre que na parte esquerda dunha unidade gráfica amalgamada apareza por exemplo a secuencia *cónto* podemos identificala coa forma verbal da primeira singular do presente de indicativo do verbo *contar*, que estará obrigatoriamente acompañada dun ou de máis dun enclítico: *cóntoa*, *cóntoo*, *cóntolle*, *cóntoche*, *cóntovola*, etc. Este sistema obríganos tamén a introducir para as formas verbais rematadas en *-r* ou *-s* unha terminación que careza dese *-r* ou *-s*, mais que posúa a mesma etiqueta caracterizadora da terminación con ese carácter, para poder permitir o recoñecemento e xeración desas formas verbais cando se empregan coa segunda forma do artigo ou co pronome de terceira acusativo. Desta maneira a terminación *-as* cuxa

subetiqueta é pi20s (presente, indicativo, segunda, xénero non aplica, singular) posibilita a caracterización da cadea verbal *cóntas* cando aparece con calquera enclítico que non sexa a segunda forma do artigo ou o pronome de acusativo de terceira: *cóntasmo*, *cóntaslle*, *cóntasnolo*, etc. Mentres que a terminación *-a* cuxa etiqueta tamén é pi20s (presente, indicativo, segunda, xénero non aplica, singular) permite a caracterización da cadea verbal *cónta* cando aparece combinada coa segunda forma do artigo ou coa variante alomórfica *lo*, *la*, *los*, *las* do pronome de terceira acusativo: *cónta-los cartos?*, *cóntalos?*, etc.

Viamos nas contraccións que nelas xa figuraba a forma que caracterizaba a cada un dos constituíntes, por exemplo en *polo* indicábase de antemán que a forma do segundo compoñente era *lo*, a correspondente á segunda forma do artigo. Mais, como reconstruír as formas dos constituíntes verbais se non estamos ante unha listaxe pechada e polo tanto non caracterizamos en ningún sitio o conxunto amalgamado de forma verbal e enclítico? En realidade é máis sinxelo do que nun principio se podería pensar.

Partamos de que o sistema identificou e caracterizou morfoloxicamente a parte esquerda da unidade gráfica plurigramatical como unha forma verbal a partir da información que lle consta no sublexicón *clíticos*. Ten o constituínte verbal identificado e etiquetado, só lle queda reconstruír a forma. Para iso é necesario comunicar os dous lexicóns, o *principal* e o de *clíticos*, só que do *principal* necesitamos crear unha copia invertida, de xeito que nel conste primeiramente a etiqueta e o lema e, a continuación, a unidade gráfica que se corresponde con eses datos. Deste modo, o sistema acode aos dous lexicóns para completar a análise deste tipo de unidades. O lexicón de formas verbais que poden levar enclíticos axúdalle a escoller a etiqueta e lema correctos, para que logo sexa no lexicón invertido onde a partir da etiqueta e lema escolla a ortografía da forma verbal correspondente. Por exemplo, tomemos a unidade *contádello*; ante ela o preprocesador, grazas ao lexicón de raíces verbais que poden levar enclíticos e aos recursos lingüísticos que forman parte do sistema e que veremos deseguida, segmenta a unidade nos constituíntes *contáde*, *lle* e *o*; logo, etiqueta *contáde* como segunda persoa do plural do imperativo do lema *contar* e, finalmente, acode ao lexicón invertido para comprobar que forma ortográfica lle corresponde a esa etiqueta e ese lema, ve cal é e substitúe a que tiña por *contade*, sen acento gráfico loxicamente. O mesmo proceso ca nos cambios acentuais prodúcese na recuperación do *-r* ou do *-s* das formas verbais que

rematan nesas consoantes cando con elas se utiliza a segunda forma do artigo ou o alomorfo *lo, la, los, las* do pronome de terceira acusativo: *contalo* <*contar + lo*; *cóntalo* <*contas + lo*, *contádelo* <*contades + lo*; etc. Por suposto, tamén se aplica o mesmo método na reconstrución das primeiras persoas do plural cando levan enclítico o pronome *nos*: *sentámonos* <*sentamos + nos*; *divertímonos* <*divertimos + nos*; etc.

2.4.2.3. Secuencias de enclíticos válidas

En segundo lugar, outro recurso imprescindible en XIADA para ofrecernos as análises que anticipamos é que o sistema conte cunha táboa, accesible a el mais non integrada no lexicón, con todas as combinacións válidas de pronomes enclíticos. Por iso elaboramos unha listaxe que contén 189 secuencias de clíticos, na que, para cada secuencia, indicamos cun valor numérico, do 1 ao 4, a lonxitude que posúe a cadea; é dicir, informamos do número de pronomes enclíticos do que consta cada secuencia. Para a elaboración desta relación de combinacións de pronomes átonos en énclice seguimos principalmente Álvarez, Regueira e Monteagudo (1986), así como Álvarez (1994). A efectos de recoñecemento e caracterización introducimos nesta relación as expresións da segunda forma do artigo, tanto que aparecen como únicos elementos en énclice como que figurasen en combinación con algún pronome átono (con *nos, vos e lles*), porque o seu comportamento computacional é idéntico ao dos pronomes átonos enclíticos. Reproducimos a listaxe completa no Anexo E por considerar que pode ser de interese mais, para que poida comprobarse xa aquí a información descrita, recollemos a modo de exemplificación as vinte primeiras secuencias:

<i>Secuencia</i>	<i>Lonxitude</i>
-lo	1
-la	1
-los	1
-las	1
o	1
a	1
os	1
as	1
lo	1
la	1

<i>Secuencia</i>	<i>Lonxitude</i>
los	1
las	1
chello	3
chella	3
chellos	3
chellas	3
che	1
chelleme	3
chelleno-lo	4
chelleno-la	4

A presenza do valor numérico ao lado da secuencia permítenos, na xeración de formas, decidir con cantos pronomes enclíticos queremos extraer as formas verbais caracterizadas en XIADA. Deste xeito énos posible producir todas as entradas do noso lexicón especificando nos verbos o número de pronomes enclíticos cos que queremos que se xeren as formas verbais.

A versión 2.4 do CORGAetq confirmanos que o normal en galego na linguaxe xornalística é a presenza dun único elemento enclítico, pois documentamos 3335 ocorrencias de forma verbal cun enclítico e 855 con segunda forma do artigo. Pola contra, o sistema non ofrece ningún resultado se buscamos tres elementos amalgamados a unha forma verbal, sexa o terceiro constituínte pronome ou artigo, mentres que a presenza de formas verbais con dous pronomes enclíticos increméntase ata contar 35 casos, malia que a variación na secuencia de pronomes se reduce a oito combinacións distintas das cincuenta e oito en teoría posibles:

selle: 18 casos; exs.: *fóiselle, permíteselle*.

llo: 6 casos; exs.: *comunicarllo, púxenllo*.

senos: 4 casos; exs.: *amósasenos, rompérsenos*.

selles: 3 casos; exs.: *cortáronselles, explicaselles*.

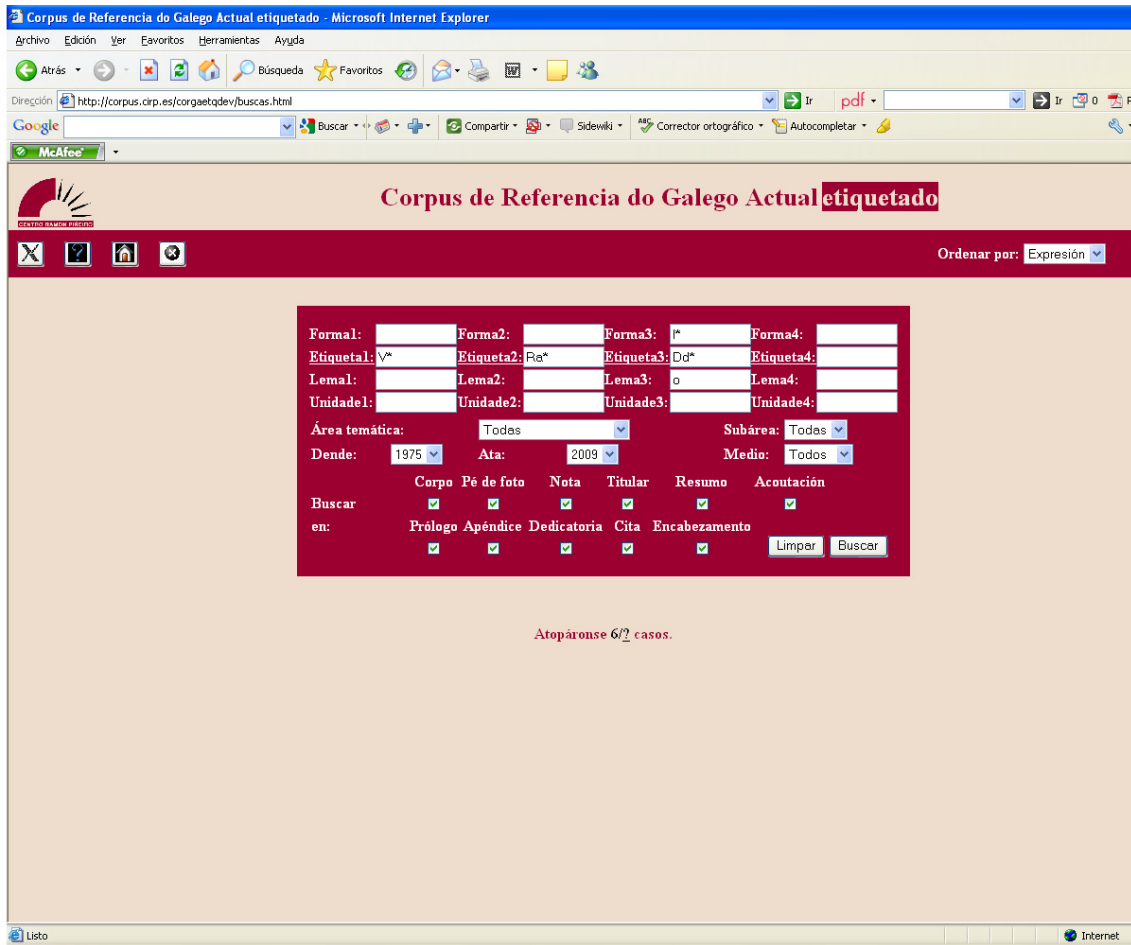
lla: 1 caso; ex.: *administrarlla*.

llelos: 1 caso; ex.: *facilitárllelos*.

mos: 1 caso; ex.: *sinaládomos*.

nola: 1 caso; ex.: *achéganola*.

Se o segundo dos constituíntes amalgamados é a segunda forma do artigo, os resultados límitanse a 6 ocorrencias:



Deses 6 casos, 5 están constituídos pola combinación do pronome de dativo de terceira plural coa segunda forma do artigo feminino –en catro ocasións este último ten número singular e noutra é plural–, namentres que na outra combinación concorre o pronome de dativo primeira plural coa segunda forma do artigo masculino plural: *dárlle-la volta, dárlle-la razón, facilitárlle-la confección, vendérlle-la totalidade; entregarálle-las insignias e dárnno-los dous exercicios.*

2.4.2.4. Secuencias clíticas monosilábicas

En terceiro lugar, o sistema debe poder acceder a unha táboa na que consten os clíticos ou combinacións de clíticos que ocupen unha soa sílaba, ou sexa, a listaxe que todos coñecemos dos pronomes átonos que poden ir enclíticos e as contraccións dos pronomes de dativo *me, che* e *lle* cos de terceira acusativo. Nela inclúense, naturalmente, as catro expresións da segunda forma do artigo:

<i>Identificador</i>	<i>Secuencia</i>	<i>Normativa2003</i>
1	-lo	si
2	-la	si
3	-los	si
4	-las	si
5	o	si
6	a	si
7	os	si
8	as	si
9	che	si
10	cho	si
11	cha	si
12	chos	si
13	chas	si
14	lo	si
15	la	si
16	los	si
17	las	si
18	lle	si
19	llo	si

20	lla	si
21	llos	si
22	llas	si
27	me	si
28	mo	si
29	ma	si
30	mos	si
31	mas	si
32	no	si
33	na	si
34	nos	si
35	nas	si
36	te	si
37	se	si
38	vo	si
39	vos	si
23	lles	si

En total conforman unha relación de 36 secuencias ás que se lles proporciona un identificador e se indica se son normativas ou non. Todas estas secuencias son consideradas normativas, porque neste nivel non se teñen en conta as etiquetas dos compoñentes. Falabamos noutro apartado da presenza nos textos do galego actual das contraccións *llo, lla, llos, llas* nas que o primeiro clítico expresa nuns casos número singular e noutros plural; pois ben, nesta táboa de enclíticos ou combinacións de enclíticos monosilábicas as secuencias *lle, llo, lla, llos, llas* son normativas porque a secuencia en si é correcta e será na descomposición da unidade en compoñentes cando, ao desagregarse as análises, se indique que esas secuencias non son normativas se o clítico *lle* se etiqueta como Rad3ap (*Pronome, átono, dativo, terceira, masculino/feminino, plural*).

Naturalmente, os datos da táboa anterior non son suficientes para achegarnos na saída as análises que desexamos, polo que debemos incrementalos, xa que o sistema non pode proporcionarnos unha información da que carece. Por iso, logo, os datos desta

táboa de clíticos monosilábicos relaciónanse con outra táboa na que se recolle, para estas mesmas secuencias monosilábicas, a seguinte información:

- o identificador da secuencia. Por exemplo, o de *llo* é o 19.
- o identificador do compoñente clítico único ou os identificadores dos compoñentes da secuencia se esta é unha contracción. Por exemplo, *llo* presenta dous compoñentes: *lle* co identificador 34 e *o* co identificador 37.
- a posición de cada compoñente. Seguindo con *llo*, indicamos que o compoñente *lle* ocupa a primeira posición e o compoñente *o* a segunda.
- o número de descomposición por se hai máis dunha análise. Para *llo* facemos constar dúas análises, polo tanto un 2 neste apartado.
- as etiquetas e lemas posibles dos distintos clíticos. O compoñente 34, *lle*, presenta dúas posibles etiquetas: unha, como Rad3as (*Pronome, átono, dativo, terceira, masculino/feminino, singular*) e outra como Rad3ap (*Pronome, átono, dativo, terceira, masculino/feminino, plural*), aínda que en realidade deberíamos falar de seis etiquetas, pois en cada un dos casos anteriores o valor ambiguo do xénero é desdoblado na análise real en tres posibilidades (*masculino/feminino; masculino; feminino*); en ambos os casos o lema é *lle*. Para o compoñente 37 figura a etiqueta Raa3ms (*Pronome, átono, acusativo, terceira, masculino, singular*) cuxo lema é *o*.
- a normatividade do compoñente. Aquí, para rematar con *llo*, é onde se lle indica que na segunda descomposición e coa etiqueta Rad3ap, o compoñente *lle* é non normativo, mentres que o compoñente *o* é normativo.

É comprensible preguntarse por que difire o comportamento entre as secuencias de clíticos monosilábicas e as polisilábicas, mellor dito, por que non incluímos as segundas nesta táboa. Tal vez incluso pensemos en que a explicación radique na acentuación da forma verbal, pois somos conscientes de que con só engadir ao verbo enclíticos de máis dunha sílaba este obrigatoriamente debe acentuarse. Pero non, a resposta é outra.

A táboa que estamos describindo só integra as secuencias monosilábicas por dous motivos. Un, o preprocesador vai sílaba a sílaba na segmentación da parte dereita

dunha unidade gráfica composta por forma verbal con enclíticos, polo que precisa coñecer que clíticos ou combinacións de clíticos forman unha sílaba; segundo, a información que necesita coñecer o sistema sobre que etiquetas pode levar cada un dos clíticos considerados como compoñentes xa aparece nestas secuencias. Debemos ver esta lista como o núcleo dos compoñentes clíticos, pois calquera que sexa a combinación de pronomes átonos enclíticos que se nos ocorra, necesariamente, ten que establecerse cos compoñentes que aparecen nesta táboa. É verdade que nela non aparecen *llelo*, *lle-lo*, *nolo*, *no-lo* nin *vollo*, *vo-lo* cos respectivos femininos e plurais, mais si están os dous compoñentes de calquera das cadeas anteriores. É probable que ao pasar a vista pola listaxe e ver *lle* ou *no* só pensáramos en que a etiqueta que lles corresponde é respectivamente a de pronome átono dativo, terceira, masculino/feminino singular (*lle*) e pronome átono, terceira, acusativo, masculino, singular (*no*). Non obstante, na antepenúltima posición e co identificador 38 aparece a secuencia *vo*, a cal non podemos identificar con ningún clítico non sendo como variante de *vos* cando vai seguido da segunda forma do artigo ou do alomorfo do pronome de terceira acusativo: *vo-lo*, *vo-la*, *vo-los*, *vo-las* ou *vollo*, *vola*, *volos*, *volas*, co que xa nos damos conta de que é probable que *lle* ou *no* presenten ademais as etiquetas de pronomes átonos dativos correspondentes, respectivamente, á terceira e á primeira de plural cando se combinan coa segunda forma do artigo ou co alomorfo do acusativo de terceira: *llelo*, *lle-lo*, *nolo*, *no-lo*, etc. E así é; para *lle*, por exemplo, constan as seguintes etiquetas:

Rad3ap [*Pronome, átono, dativo, terceira, masculino/feminino, plural*]

Rad3as [*Pronome, átono, dativo, terceira, masculino/feminino, singular*]

Rad3fp [*Pronome, átono, dativo, terceira, feminino, plural*]

Rad3fs [*Pronome, átono, dativo, terceira, feminino, singular*]

Rad3mp [*Pronome, átono, dativo, terceira, masculino, plural*]

Rad3ms [*Pronome, átono, dativo, terceira, masculino, singular*]

Deste xeito, o preprocesador cando se atopa cunha secuencia de clíticos, primeiro comproba que estea na táboa que inclúe todas as combinacións posibles. A continuación mira se a secuencia conforma unha sílaba ou máis dunha sílaba; se constitúe unha soa sílaba acode á táboa de secuencias monosilábicas e comproba toda a información que se relaciona con esa secuencia para achegarnos todas as análises

posibles: segmentación en constituíntes, se posúe máis dun compoñente, e todas as etiquetas con eles asociadas. Cando a secuencia de clíticos conforma máis dunha sílaba o preprocesador primeiro vana descompoñer en sílabas e logo consulta a táboa de secuencias monosilábicas para comprobar que caracterización morfosintáctica lle corresponde a cada un dos constituíntes. En ambos os casos testará as regras que veremos a continuación para que lle axuden na segmentación e caracterización.

Non aproveitamos para os pronomes átonos a información que posuímos nas contraccións sobre os distintos compoñentes, porque nelas non están recollidas todas as contraccións de átonos que poden ir enclíticas a unha forma verbal. Así, as combinacións de *nos*, *vos* e *lles* coas segundas formas do artigo unicamente poden aparecer engadidas a un verbo; ademais, nalgún lugar temos que individualizar e caracterizar os pronomes simples que poden ir enclíticos, porque o lexicón xeral só acolle as entradas que poden aparecer nun texto entre dous espazos en branco co que, computacionalmente, nos crea máis problemas separar as ocorrencias que deben caracterizarse nas contraccións das demais ca crear unha táboa única coas secuencias monosilábicas e as súas características.

Tal como explicamos, non necesitamos almacenar todas as etiquetas das combinacións de clíticos, senón só as básicas. Ademais, en caso de querer almacenalas teríamos que gardar a información para todas e cada unha das combinacións de clíticos, motivo polo que o descartamos ao considerar que era unha tarefa excesiva, xa que podíamos lograr o mesmo obxectivo almacenando as etiquetas de cada un dos compoñentes e definindo unhas poucas regras que eliminaran as combinacións imposibles.

2.4.2.5. Regras lingüísticas

Antes de continuar debemos puntualizar que nós non nos ocupamos de implementar as regras lingüísticas que gobernan o funcionamento do preprocesador, senón simplemente de enumeralas e describilas. Trátase dun conxunto de regras, elaboradas en formato xml, para axudarlle ao sistema tanto na segmentación e etiquetación de enclíticos como na das formas verbais cando van acompañadas de enclíticos.

Podemos dividilas en dous subconxuntos, un aplicado única e exclusivamente sobre uns poucos clíticos e outro que incide tanto na segmentación como na etiquetación das formas verbais e os clíticos que as acompañan.

2.4.2.5.1. Regras de clíticos

A diferenza do que ocorre nos dous sistemas do SLI da Universidade de Vigo que se aplican na etiquetación de corpus galegos, en XIADA a presenza dun clítico nunha contracción condiciona a súa etiquetación.

Na etiquetación de pronomes, tanto no TECTRA como no CTAG, caracterízanse os clíticos sen ter en conta que polo mero feito de figurar nunha contracción ou nunha combinación determinada se restrinxen xa as posibilidades de análise. Así, por exemplo, no sistema empregado polo SLI-TECTRA indícase que na secuencia *mo* o último pronome é “pronome persoal átono masculino singular terceira acusativo”, pero para o primeiro non se di cal é o seu caso, só se indica que é un “pronome persoal átono singular primeira” (Aguirre Moreno et al., 2002b: 224):

díxomo: VIPES3_PPS1A_PPMS3AA [“verbo indicativo pretérito terceira singular” + “pronome persoal átono singular primeira” + “pronome persoal átono masculino singular terceira acusativo”]

No sistema do CTAG recóllese o caso na descrición do clítico, mais este non se especifica. Así no exemplo seguinte (Gómez Guinovart e López Fernández, 2009: 69), a pesar de que a combinación co artigo indica claramente que o enclítico é dativo, mantense na etiqueta o caso “acusativo/dativo/reflexivo”:

dóuvo-la/dar_VIP1S00 ~/vos_P20PC0 ~/o_GFS [“verbo indicativo presente primeira singular”; “pronome segunda ¿xénero non especificado?¹¹⁸ plural acusativo/dativo/reflexivo cortésia non especificada”; “determinante feminino singular”]

Naturalmente, en ambos os sistemas as descricións veñen condicionadas pola elección do etiquetario no que non se teñen en conta os casos concretos, senón que se realiza unha caracterización xenérica.

¹¹⁸ Segundo a táboa de atributos e valores e mais os exemplos (Gómez Guinovart e López Fernández, 2009: 65) nesta etiqueta para expresar o valor do xénero debería aparecer “C” (común), mais o que en realidade figura é “0”.

En XIADA, polo contrario, o etiquetario ofrece a posibilidade de que os pronomes átonos poidan caracterizarse atendendo a cada caso concreto, de aí que ante unha secuencia como *nola* presente por exemplo na seguinte oración da versión 2.4 do CORGAetq

Unha das estreas máis bizarras do ano achéganola un dos cineastas máis estrambóticos dos últimos anos, Harmony Korine. [CORGAetq: ANT2008-07-10/66]

cando o preprocesador identifica *achéga* como unha forma verbal con clíticos, que segundo o lexicón de raíces verbais susceptibles de construírse con clíticos pode etiquetarse como V0m20s (*Verbo, tempo non aplica, imperativo, segunda, xénero non aplica, singular*) ou Vpi30s (*Verbo, presente, indicativo, terceira, xénero non aplica, singular*), e a continuación comproba na táboa de secuencias de clíticos válidas que a combinación *nola* está presente e posúe dous compoñentes, daquela comproba as posibles etiquetas para cada un dos compoñente e atópase con que o primeiro deles posúe varias posibilidades de análise:

Rad1ap	[Pronome, átono, dativo, primeira, masculino/feminino, plural]
Rad1fp	[Pronome, átono, dativo, primeira, feminino, plural]
Rad1mp	[Pronome, átono, dativo, primeira, masculino, plural]
Raa1ap	[Pronome, átono, acusativo, primeira, masculino/feminino, plural]
Raa1fp	[Pronome, átono, acusativo, primeira, feminino, plural]
Raa1mp	[Pronome, átono, acusativo, primeira, masculino, plural]

Entre as posibles caracterizacións de *no* non está incluída a de alomorfo do pronome de terceira acusativo masculino singular (lema *o*) porque, como veremos no punto seguinte, as regras que rexen a combinación de formas verbais con pronomes enclíticos indican que a parte verbal non remata en ditongo, co que esa posible análise queda invalidada.

O sistema accede entón ás regras de enclíticos e verifica as restricións que nelas aparecen para cada caso concreto e conclúe que a etiqueta que lle corresponde é a de Rad1ap, Rad1fp ou Rad1mp pois atopou unha regra que o informa de que se ao enclítico *no* o segue o enclítico *lo, la, los* ou *las* ten que filtrar/eliminar as etiquetas correspondentes ao caso acusativo. A regra que está decidindo en realidade é a seguinte:

```

<rule>
<enclitic>no</enclitic>
<condition>
  <target>complete_form</target>
  <content>
    <evaluation at="next_enclitic">lo OR la OR los OR las OR -lo OR -la OR -los OR -
    las</evaluation>
  </content>
  <action>filter</action>
  <check_default>no</check_default>
  <filter param="nos">replace</filter>
  <filter>Raa1ap OR Raa1fp OR Raa1mp</filter>
</condition>
</rule>

```

As regras de clíticos implementadas en XIADA afectan ás secuencias pronominais máis problemáticas computacionalmente: *lle*, *no* e *nos*. Os problemas no seu recoñecemento xorden porque calquera desas tres secuencias presenta máis dunha análise posible. *Lle*, porque se confunde na expresión como clítico singular (*déalleme isto a Xoán, por favor*) coa forma que aparece en combinación coa segunda forma do artigo ou co alomorfo do pronome de terceira acusativo (*os xoguetes déallos aos nenos*); e *no* e *nos* polo mesmo motivo, pois formalmente coinciden na expresión o alomorfo *no* do pronome de terceira acusativo (lema *o*) coa variante de *nos*, *no* (lema *nos*), cando vai seguida da segunda forma do artigo ou do alomorfo do pronome de acusativo.

As regras creámolos para habilitar unha maneira de axudarlle ao preprocesador e ao etiquetador a decidir a segmentación e etiquetación nestes casos nos que se producían erros recorrentes. Non son moi numerosas e céntranse, sobre todo, en aclarar cando as formas *no* e *nos* se corresponden co pronome de terceira e cando co de primeira plural e cando *lle* é singular e cando plural. En cada un dos casos tense en conta se o enclítico é o último da serie ou se ocupa unha posición intermedia, e en función diso fórmulase a regra. Non as imos ver aquí todas, mais si nos gustaría recoller as que lle afectan a *lle*, por ser as menos¹¹⁹.

Dúas son as regras implementadas actualmente en XIADA para *lle*, que nun caso ocupa a posición intermedia (*intermediate_enclitic*) e noutro a final (*final_enclitic*). Na

¹¹⁹ Polo momento non se modificaron as regras existentes nin se implementaron máis para axudar na diferenciación do *lle* normativo do *lle* plural non normativo. A pesar de que nas contraccións e nas secuencias de clíticos consultadas polo sistema figuran xa as análises normativas e non normativas de *llo(s)*, *lla(s)* aínda temos que ver de que maneira podemos implementar a nova regra para permitir a análise do *lle* non normativo como Rad3ap, pois esta deberá ter en conta o contexto.

primeira das regras establécense dúas condicións e o filtro aplícase sobre etiquetas distintas dependendo de cada condición. Na primeira das condicións dispónse que se o seguinte enclítico é *lo, la, los, las, -lo, -la, -los, ou -las* non se lle asignen a *lle* etiquetas co número singular, mentres que se o seguinte enclítico non é ningún dos anteriores, segunda condición, establécese que as etiquetas a asignar non poidan ser as que leven número plural. Finalmente, a última regra indica que se *lle* aparece ao final da secuencia de enclíticos non pode ser plural.

As regras para *lle*, tal e como foron formuladas, son as seguintes:

```

<intermediate_enclitic>
  <rule>
    <enclitic>lle</enclitic>
    <condition>
      <target>complete_form</target>
      <content>
        <evaluation at="next_enclitic">lo OR la OR los OR las OR -lo OR -la
          OR -los OR -las</evaluation>
      </content>
      <action>filter</action>
      <check_default>no</check_default>
      <filter param="lles">replace</filter>
      <filter>Rad3as OR Rad3ms OR Rad3fs</filter>
    </condition>
    <condition>
      <target>complete_form</target>
      <content>
        <evaluation at="next_enclitic">NOT lo AND NOT la AND NOT los
          AND NOT las AND NOT -lo AND NOT -la AND NOT -los AND
          NOT -las</evaluation>
      </content>
      <action>filter</action>
      <check_default>no</check_default>
      <filter>Rad3ap OR Rad3mp OR Rad3fp</filter>
    </condition>
  </rule>
</intermediate_enclitic>
<final_enclitic>
  <rule>
    <enclitic>lle</enclitic>
    <condition>
      <target>verb_part</target>
      <content>
        <evaluation at="end"></evaluation>
      </content>
      <action>filter</action>
      <check_default>no</check_default>
      <filter>Rad3ap OR Rad3mp OR Rad3fp</filter>
    </condition>
  </rule>
</final_enclitic>

```

2.4.2.5.2. Regras de verbos con enclíticos

O sistema principal de XIADA acode ao subsistema verbal, o formado polo lexicón de formas verbais susceptibles de levar enclíticos e mais o conxunto de regras elaboradas para as formas verbais con enclíticos, sempre que se atopa cunha palabra cuxa parte inicial aparece no lexicón de raíces verbais con clíticos e delega nel a segmentación e etiquetación da parte verbal. Nalgúns casos este proceso pode realizarse varias veces para o mesmo *token* posto que pode haber varias posibilidades de segmentación para a mesma unidade.

A estrutura das regras verbais comprende as seguintes partes:

A) Zona de aplicación ou *target*. Refírese á parte da unidade gráfica sobre a que vai incidir a regra. Pode ser a parte verbal (*verb_part*), a forma completa (*complete_form*) ou as etiquetas verbais (*verb_tag*).

B) Contido da regra ou *content*. Expresa a condición da regra e contén un ou máis elementos de avaliación coas correspondentes expresións.

C) Acción a executar ou *action*. Determina a acción a executar se se cumpre a condición expresada. Pode indicar *continuar*, se segue a avaliación na seguinte expresión; *aceptar*, cando remata a avaliación confirmando que efectivamente se trata dunha segmentación correcta; *rexear*, cando remata a avaliación confirmando que non se trata dunha segmentación correcta ou *filtrar*, que coincide con *continuar* pero elimina as etiquetas especificadas nos elementos *filter*.

D) *Check_defaul*. Determina se a regra establecida por defecto debe ser avaliada ou non; é dicir, indica se se debe ter en conta a regra principal do sistema, aquela que di que se a parte verbal remata en vogal ou *-n* o primeiro clítico non pode ser o alomorfo de terceira acusativo *lo* nin a segunda forma do artigo.

E) *Filter*. Especifica o conxunto de etiquetas que deben eliminarse antes de continuar a regra. Contén unha expresión booleana que verifica todas as etiquetas da parte verbal. As etiquetas que se axustan á expresión booleana elimínanse e o sistema continúa coas comparacións. Isto é moi útil cando existen etiquetas non válidas para a segmentación en curso.

O subsistema comproba as regras dende a primeira ata a última. Se a expresión dunha condición corresponde á parte da unidade sinalada, execútase a acción marcada. O proceso cesa cando na acción topa cun *aceptar* ou *rexear* ou despois de comprobar todas as regras. Se non hai unha regra que rexeite a segmentación, o sistema acéptaa por defecto. Malia que non imos ver aquí o sistema completo das regras verbais con clíticos, pois este tema podería ser obxecto doutra tese, si describiremos basicamente o seu contido e funcionalidade.

As regras de formas verbais con enclíticos, elaboradas para especificar que partes esquerdas (formas verbais) poden combinarse con que partes dereitas (secuencias de clíticos), débense a que sen elas calquera raíz verbal do sublexicón *clíticos* podería combinarse con calquera secuencia de enclíticos, e o sistema non sabería discriminar se as segmentacións e etiquetacións son correctas ou incorrectas. Aplícanse principalmente en dous casos:

1) Para indicarlle ao preprocesador que determinadas raíces verbais caracterizadas con certas etiquetas deben levar máis dun elemento enclítico para ser análises válidas.

Tal e como está establecido o sistema de XIADA énos máis cómodo incidir coas regras en que raíces non poden levar un só clítico, pois tanto as terminacións agudas como as secuencias de clíticos monosilábicas coas que non poden combinarse constitúen unha cifra menor, ca enumerar as raíces tónicas e clíticos cos que estas poden combinarse. Deste xeito non só damos conta computacionalmente das repercusións que na acentuación da forma verbal ten a éncise de pronomes, senón que tamén filtramos combinacións incorrectas. Por exemplo, o preprocesador debe rexeitar a segmentación ou etiquetación de calquera forma verbal rematada en *-ás* ou *-á*, cuxas etiquetas sexan Vpi20s, V0m20s ou Vpi30s, se esta leva un único enclítico, exceptuando por suposto as formas *dás* e *dá*:

```
<rule>
  <condition>
    <target>verb_part</target>
    <content>
      <evaluation at="end">ás OR á</evaluation>
    </content>
    <action>continue</action>
    <check_default>yes</check_default>
    <filter></filter>
```

```

</condition>
<resolution>
<condition>
    <target>verb_part</target>
    <content>
        <evaluation at="all">NOT dás NOT dá</evaluation>
    </content>
    <action>continue</action>
    <check_default>yes</check_default>
    <filter></filter>
</condition>
<condition>
    <target>tag</target>
    <content>
        <evaluation at="all">Vpi20s OR Vpi30s OR
V0m20s</evaluation>
    </content>
    <action>filter</action>
    <check_default>yes</check_default>
    <filter value="1">count</filter>
</condition>
</resolution>
</rule>

```

Outro exemplo de regra que restrinxe as raíces verbais en combinación cun só clítico é a que establece que se a parte verbal remata en *-ín* e a etiqueta que lle corresponde é a de *Vei10s* (*Verbo, pretérito, indicativo, primeira, xénero non aplica, singular*), se leva un único enclítico, rexeite esa caracterización a non ser que a forma verbal sexa algunha das que están incluídas nos grupos *Vic4a*, *Vic5b* ou *Vic7a*:

```

<rule>
    <condition>
        <target>verb_part</target>
        <content>
            <evaluation at="end">ín</evaluation>
        </content>
        <action>continue</action>
        <check_default>yes</check_default>
        <filter></filter>
    </condition>
    <resolution>
    <!-- Excepcións grupo Vic4a -->
        <condition>
            <target>verb_part</target>
            <content>
                <evaluation at="all">NOT abstraín AND NOT acain AND
                NOT atraín AND NOT caín AND NOT choín AND NOT
                condoín AND NOT contraín AND NOT corroín AND NOT
                debaín AND NOT decaín AND NOT degraín AND NOT
                deschoín AND NOT descontraín AND NOT detraín AND
                NOT distraín AND NOT doín AND NOT esvaín AND NOT
                extraín AND NOT maltraín AND NOT moín AND NOT
                proín AND NOT raín AND NOT recaín AND NOT remoín
                AND NOT retraín AND NOT retrotraín AND NOT roín
            </content>
        </condition>
    </resolution>
</rule>

```

```

                AND NOT substraín AND NOT subtraín AND NOT
                sustraín</evaluation>
            </content>
            <action>continue</action>
            <check_default>yes</check_default>
            <filter></filter>
        </condition>
    <!-- Excepcións grupo Vic5b -->
    <condition>
        <target>verb_part</target>
        <content>
            <evaluation at="all">NOT desoín AND NOT entreoín AND
            NOT esvaín AND NOT oín AND NOT saín AND NOT
            sobresaín</evaluation>
        </content>
        <action>continue</action>
        <check_default>yes</check_default>
        <filter></filter>
    </condition>
    <!-- Excepcións grupo Vic7a -->
    <condition>
        <target>verb_part</target>
        <content>
            <evaluation at="all">NOT abluín AND NOT afluíñ AND
            NOT arguíñ AND NOT atribuíñ AND NOT atuín AND NOT
            concluín AND NOT confluín AND NOT constituín AND
            NOT construíñ AND NOT contribuíñ AND NOT derruíñ
            AND NOT desatuín AND NOT desobstruíñ AND NOT
            desposuíñ AND NOT destituín AND NOT destruín AND
            NOT difluín AND NOT diluíñ AND NOT diminuíñ AND
            NOT disminuíñ AND NOT distribuín AND NOT efluín
            AND NOT esluín AND NOT estatuín AND NOT excluín
            AND NOT fluín AND NOT imbuín AND NOT incluín AND
            NOT influín AND NOT immiscuín AND NOT instituín AND
            NOT instruín AND NOT intuín AND NOT luín AND NOT
            obstruíñ AND NOT ocluín AND NOT posuíñ AND NOT
            prostituín AND NOT puín AND NOT recluíñ AND NOT
            reconstituín AND NOT reconstruíñ AND NOT redistribuíñ
            AND NOT refluíñ AND NOT restituín AND NOT retribuín
            AND NOT substituíñ AND NOT sustituíñ</evaluation>
        </content>
        <action>continue</action>
        <check_default>yes</check_default>
        <filter></filter>
    </condition>
    <condition>
        <target>tag</target>
        <content>
            <evaluation at="all">Vei10s</evaluation>
        </content>
        <action>filter</action>
        <check_default>yes</check_default>
        <filter value="1">count</filter>
    </condition>
</resolution>
</rule>

```

2) No segundo caso as regras céntranse en indicar non cantos clíticos poden acompañar ás formas verbais senón en determinar que clíticos non poden combinarse cunhas determinadas formas verbais. Así, por exemplo, hai unha regra que lle impide ao preprocesador dar por válida a análise dunha forma verbal rematada en ditongo se como primeiro elemento enclítico desta aparecen os pronomes *o* ou *lo*, con calquera das súas formas, ou a segunda forma do artigo:

```
<rule>
  <condition>
    <target>verb_part</target>
    <content>
      <evaluation at="end">ei OR éi OR eu OR éu OR ou OR óu OR iu OR
      íu OR ai OR ái OR oi OR ói</evaluation>
    </content>
    <action>continue</action>
    <check_default>no</check_default>
    <filter></filter>
  </condition>
  <resolution>
    <condition>
      <target>complete_form</target>
      <content>
        <evaluation at="first_enclitic">o OR os OR a OR as OR la OR las OR
        lo OR los OR -lo OR -la OR -los OR -las</evaluation>
      </content>
      <action>reject</action>
      <check_default>no</check_default>
      <filter></filter>
    </condition>
  </resolution>
</rule>
```

Outra mostra deste tipo de regras constitúea a establecida para que o pronome *se* só poida combinarse con formas verbais de terceira persoa, singular ou plural, de todos os tempos verbais persoais e co infinitivo e o xerundio:

```
<rule>
  <condition>
    <target>complete_form</target>
    <content>
      <evaluation at="first_enclitic">se</evaluation>
    </content>
    <action>filter</action>
    <check_default>yes</check_default>
    <filter>NOT V0f000 AND NOT V0x000 AND NOT V??30?</filter>
  </condition>
</rule>
```

Con todo, esta regra precisa doutra para rexeitar casos coma **cantásese*, **comésese*, etc., porque o pronome *se* non concorre nunca cunha forma verbal rematada

en *-se* como son as formas do pretérito de subxuntivo. A existencia de *cósesese* implica que teñamos que recorrer ás etiquetas para concretar a regra e que non poidamos limitarnos a expresala atendendo unicamente ás terminacións verbais:

```

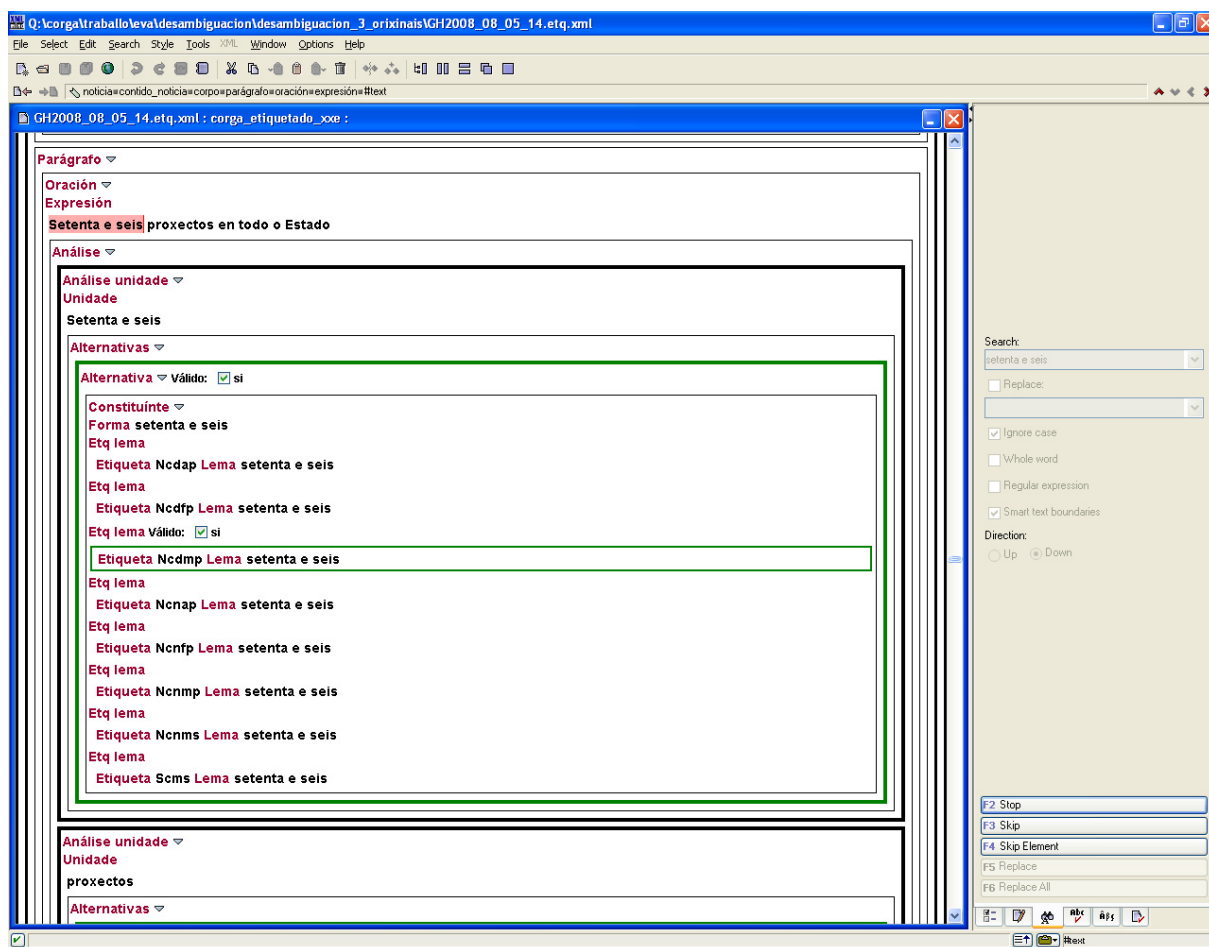
<rule>
  <condition>
    <target>verb_part</target>
    <content>
      <evaluation at="end">se</evaluation>
    </content>
    <action>continue</action>
    <check_default>yes</check_default>
    <filter></filter>
  </condition>
  <resolution>
    <condition>
      <target>tag</target>
      <content>
        <evaluation at="all">Ves30s</evaluation>
      </content>
      <action>continue</action>
      <check_default>yes</check_default>
      <filter></filter>
    </condition>
    <condition>
      <target>complete_form</target>
      <content>
        <evaluation at="first_enclitic">se</evaluation>
      </content>
      <action>reject</action>
      <check_default>yes</check_default>
      <filter></filter>
    </condition>
  </resolution>
</rule>

```

2.5. Os numerais multipalabra

Pouco é o que podemos explicar polo momento sobre o tratamento deste frecuente e ilimitado grupo de unidades multipalabra, pois forma parte dun pequeno conxunto de tarefas que fomos pospoñendo, sobre todo, por falta de tempo. Temos claro que a maneira de posibilitar o seu recoñecemento e caracterización non é a través do lexicón, senón do establecemento de regras que permitan identificalos. A principal, xa funcionando en XIADA, establece que se nos atopamos cun numeral cardinal, dos introducidos no lexicón, seguido da conxunción copulativa *e* e doutro numeral, se trate o conxunto formado por $N^1 + e + N^2$ como un único numeral cardinal. Así, na versión orixinal etiquetada do documento do CORGAetq GH2008_08_05_14.etq.xml, sen

desambiguar manualmente, podemos comprobar como o sistema identifica dous numerais cardinais consecutivos como unha única unidade:



A esta regra debe sumarse, como mínimo, outra variante na que se estableza que se aparece un numeral seguido doutro numeral, sexan cardinais sexan ordinais, se considere unha única unidade mais, polo momento, non está implementada co que o sistema non considera unidades multipalabra, por exemplo, *vixésimo quinta* ou *tres mil*, presentes respectivamente nas oracións seguintes, e ofrece análises individuais para cada unha das formas gráficas:

Tres mil soldados son despregados nas principais cidades do país como parte dunha nova medida de Berlusconi para combater a delincuencia e mellorar a orde [CORGAetq: GH200808-05/64]

Tres mil sinaturas para trasladar ó Concello [CORGAetq: GH2007-03-17/6]

A vixésimo quinta edición da Mostra de Teatro Cómico e Festivo de Cangas aposta por cubrir diferentes fronteiras da creación contemporánea mesturando espectáculos de moi distinta natureza. [CORGAetq: ANT2008-07-10/24]

Con todo, estas dúas regras non abondan para que o sistema identifique todos os numerais nin, polo de pronto, cobren todas as posibles análises. Pensemos por exemplo na primeira das regras: $N^1 + e + N^2 = N$. Sempre que nos atopamos con dous numerais seguidos podemos afirmar que forman parte da mesma unidade? A resposta ten que ser non necesariamente. É probable que si o sexan, pero tamén é posible que non. Por exemplo, ante unha secuencia como *Comprei vinte e cinco xa os comín* o sistema actúa mal, porque a regra que condiciona a segmentación e escolla da etiqueta debe conter excepcións. Precisamos, pois, estudar a gramática formal dos numerais e a partir dela establecer mediante regras as pautas que determinen a súa segmentación e etiquetación.

2.6. Datos xerais

Implementamos o lexicón coas entradas do *Vocabulario Ortográfico da Lingua Galega* (VOLGa) e coas 100 000 formas máis frecuentes do CORGA, o que se traduce en algo máis de 55 000 lemas.

Como cremos que é interesante coñecer de que xeito se reparten eses lemas entre as distintas categorías gramaticais amosamos a súa distribución nunha táboa. Non obstante, tamén damos conta do número de raíces que se introduciron no lexicón de XIADA en cada un dos grandes subtipos en que se dividen as diferentes categorías, sen esquecernos tampouco de proporcionar o número de formas que é posible xerar a partir das distintas entradas. Veremos, pois, que a información proporcionada se desagrega nalgúns casos en varias filas aínda que tamén se achegan os datos totais por categoría.

Preferimos separar os datos correspondentes á clase verbal nunha táboa independente para poder reflectir así a diferenza entre verbos regulares e irregulares e os tipos existentes dentro destes dous grandes bloques.

Antes de amosar a táboa cos diversos datos debemos realizar unhas observacións sobre estes. Nestes momentos en XIADA só podemos saber cantas entradas temos no lexicón por categoría gramatical, etiqueta/subetiqueta ou grupo de derivación, mais non podemos consultar directamente o número de lemas que existen por cada categoría ou subcategoría nin o número de formas a que dan orixe. Para cubrir ese oco informativo, en tanto non se habilitan os medios necesarios no propio sistema, recorreremos á versión 2.4 do léxico descargable de XIADA en todas as categorías, excepto para as locucións,

siglas, abreviaturas e símbolos dos que, por non estar alí incluídos, só achegamos o número de entradas que para eles temos en XIADA.

Consideramos que a información que obtemos deste léxico é reflexo da do lexicón porque foi xerada a partir da que alí consta, e así, coa excepción das categorías Locución e Categoría Periférica, o léxico descargable recolle todas as entradas que é posible xerar a partir do lexicón, 721 073 entradas nas que non aparece ningún tipo de unidade gráfica multigramatical, ou sexa, nin contraccións, nin formas verbais con enclíticos nin, por suposto, calquera unidade que para expresarse necesite apoiarse noutra, como son as segundas formas do artigo, alomorfos pronominais, etc., porque tampouco estas unidades figuran no lexicón.

Entendemos por forma a unidade ortográfica que aínda sendo idéntica a outra graficamente está asociada a unha etiqueta distinta. Por exemplo, tanto no léxico descargable como consultando por forma en XIADA aparecen tres formas *atleta*, etiquetadas todas como Adxectivo, grao non aplica singular, mais diferenciándose no valor do xénero: *masculino/feminino*, *masculino* e *feminino*. O mesmo sucede noutras categorías, coma no verbo, onde aparecerán tantas formas idénticas como etiquetas distintas lles correspondan. Isto débemolo ter en conta cando nos fixemos nas cifras da columna *Forma* mentres que na columna de *Formas únicas* recóllense, como o seu nome indica, as formas gráficas distintas que hai por cada subtipo sen ter en conta a etiqueta. Así, en tanto que na columna *Formas* se contabilizan tres formas *atleta*, na columna *Formas únicas* só se computa unha forma *atleta*.

Polo que se refire aos lemas debemos aclarar que as formas dos diminutivos, aumentativos ou superlativos sintéticos regulares remítense ao lema base, polo que non conforman un lema á parte. Por exemplo, *gravísimo*, *amodiño* ou *caladiños* están asociados respectivamente aos lemas *grave*, *amodo* e *calado* e non incrementan, en consecuencia, o cómputo de lemas. A este respecto debemos puntualizar, así mesmo, que o modo que temos de pescudar o número de lemas distintos por subtipos dentro das múltiples categorías provoca que só obteñamos un cálculo aproximado e non o exacto¹²⁰.

¹²⁰ Posto que o léxico descargable achega para cada entrada, entre outra información, a forma, etiqueta e lema, empregamos o grep para inventariar as entradas por subtipo. Así obtemos o número de formas por

<i>Subetiqueta</i>	<i>Raíces</i>	<i>Lemas</i>	<i>Formas</i>	<i>Formas únicas</i>	<i>Total lemas por categoría</i>
Sc [substantivos comúns]	30 668	30 509	75 389	68 574	30 509
Sp [substantivos propios]	0	0	0	0	
A0 [adxectivos, grao non aplica]	14 162	14 129	66 631	46 675	14 141
As [adxectivos superlativos]	13	34	140	132	
Ac [adxectivos comparativos]	7	7	42	14	
V [verbos]	9455	6903	573 653	392 735	
Wa [adverbios nuclear/modificador]	1774	0	0	0	1962
Wn [adverbios nucleares]	193	1957	1967	1967	
Wm [adverbios modificadores]	4	1774	1777	1777	
Wr [adverbios relativos]	4	4	4	4	
Wg [adverbios interrogativos]	7	4	7	7	
Dd [artigos determinados]	3	2	5	5	
Di [artigos indeterminados]	1	1	4	4	3
E [demostrativos]	9	3	30	18	
Cc [conxuncións coordinantes]	9	9	9	9	43
Cs [conxuncións subordinantes]	34	34	34	34	
G [interrogativos-exclamativos]	8	4	82	16	4
I [indefinidos]	47	38	378	116	38
La0 [locucións adverbiais]	403				
Lp0 [locucións prepositivas]	95				
Lcc [locucións conxuntivas coordinantes]	3				
Lcs [locucións conxuntivas subordinantes]	84				
M [posesivos]	11	10	84	34	10
Nc [numerais cardinais]	93	49	323	63	79
No [numerais ordinais]	30	30	240	120	
P [preposicións]	52	52	52	52	52
Q [signos de puntuación]	20	20	20	20	20
Rt [pronomes tónicos]	21	17	86	25	25
Ra [pronomes átonos]	16	8	47	12	
T [relativos]	5	5	32	12	5
Y [interxeccións]	91	91	91	91	91
Za [abreviaturas]	875				
Zg [siglas]	2172				
Zs [símbolos]	146				

Observacións:

As celas correspondentes aos substantivos propios presentan cero porque no lexicón non se introduciu ningún elemento deste tipo.

A diferenza entre o número de raíces e o de lemas nos adxectivos superlativos devén de que no lexicón só hai 13 definidos na raíz como superlativos: os de *-érrimo*

subetiqueta; para obter o lema necesitamos dar máis pasos, pois cada entrada trae o lema, repetíndose deste xeito este tantas veces como elementos flexivos posúa o mesmo paradigma. Por iso, listamos os lemas de cada unha das consultas realizadas por subtipo e eliminamos os repetidos. Así obtemos un cálculo aproximado dos lemas que constan no lexicón para esa subetiqueta. Sabemos que nos van saír menos dos que realmente hai porque, por exemplo, para os dous lemas *galo* existentes en XIADA, o do xentilicio e o do animal doméstico, ao eliminar os repetidos só aparecerá un, mais é o único método que podemos empregar nestes momentos para proporcionar datos relativos aos lemas.

(*celebérrimo, misérrimo...*), os irregulares (*pésimo, óptimo, etc.*) e os dos prefixos *arqui-* e *hiper-* (*arquimillonario, hipersensible...*). Os superlativos en *-ísimo* van ao grupo G43 e na raíz só se indica A (*Adxectivo*), polo que non os incluímos na columna das raíces, aínda que si se contabilizan nas outras columnas: lemas, formas e formas únicas.

Os adverbios en *-mente* foron introducidos no lexicón co atributo *nuclear/modificador*, pero na etiquetación ese valor non procede, co que todo adverbio ou funciona como nuclear ou como modificador; de aí que para o subtipo Wa (*Adverbio nuclear/modificador*) existan entradas no lexicón pero non se rexistren nas formas xeradas nin nos lemas.

A diferenza entre os lemas e as formas nos adverbios nucleares procede da inclusión de 10 formas co sufixo *-iño* (*amodiño, proutiño...*) *-iña* (*cerquiña, enseguidiña...*) ou o intensificador *-ísimo* (*pouquísimo, moitísimo...*), mentres que a diferenza de 3 nos adverbios modificadores entre os lemas e as formas derívase da presenza das formas *moi, moitísimo, pouquísimo e tan*. A suma total dos lemas adverbais non é igual á suma dos lemas dos adverbios nucleares (1957) e dos interrogativos e exclamativos e mais os relativos (4), porque nos adverbios modificadores hai un que non aparece nunca como nuclear: *recén*.

2.7. Datos verbais

O feito de que o 79,55% das formas do léxico descargable da versión 2.4 de XIADA sexan verbais, en concreto 573 653 formas dun total de 721 073, xustifica que creemos un apartado propio para detallar os datos da categoría Verbo.

Non debemos esquecernos de que no cómputo de formas verbais non están incluídas as posibles combinacións con enclíticos nin figuran, loxicamente, na táboa xeral as formas verbais que se poden xerar se levasen enclíticos. Aínda que o número e identidade dos lemas verbais nun caso e no outro é idéntico –6903–, o número de raíces necesarias para implementalos en XIADA, discriminando se as formas resultantes poden levar enclíticos ou non, provoca unha enorme disparidade e así, fronte ás 9455 raíces da conxugación normal, a susceptible de construírse con enclíticos case dobra esa cifra, situándose en 17 849 raíces:

<i>Sublexicón</i>	<i>Nº de raíces</i>	<i>Nº de lemas</i>
principal	9455	6903
con clíticos	17 849	

Escollemos desagregar a información relativa ao paradigma verbal, por considerar que é a máis representativa, nas seguintes columnas:

Grupo: nela figuran os 3 grupos regulares e os 42 irregulares descritos ao longo da epígrafe sobre a implementación no lexicón das categorías variables.

Número de lemas: aquí inserimos o número de lemas que se remiten a cada un dos grupos no noso sistema.

Modelos: tomamos como referencia un dos lemas que integran o grupo.

Número de etiquetas por modelo: recollemos nesta columna, como indica o seu encabezamento, o número de etiquetas que presenta cada un dos modelos cos que se exemplifica. O normal é que sexan 74, distribuídas do seguinte xeito:

Indicativo	Presente	6
	Copretérito	7
	Pretérito	6
	Pospretérito	7
	Antepretérito	7
	Futuro	6
Subxuntivo	Presente	7
	Pretérito	7
	Futuro	7
	Imperativo	2
Formas nominais	Infinitivo	1
	Xerundio	1
	Participio	4
	Infinitivo conxugado	4
	Xerundio conxugado	2

Dos números aí rexistrados, tres son os que necesitan explicación. En primeiro lugar, seis son os tempos verbais para os que proporcionamos 7 etiquetas. A pregunta que está no aire é cal é a sétima? Pois ben, todos estes tempos teñen en común compartir formalmente a primeira e a terceira persoas do singular, polo que achegamos a etiqueta que neutraliza estas dúas persoas para empregala en casos nos que non sexa posible desambiguar. En segundo lugar, pode sorprendere a inclusión do xerundio conxugado, mais o sistema necesita as etiquetas para dar resposta á realidade lingüística galega actual na que esas formas se empregan. Finalmente, para o infinitivo persoal non

recollemos as etiquetas correspondentes á primeira e á terceira persoas do singular por non ser posible a súa desambiguación co infinitivo.

Número de formas por modelo: como indica o título, introducimos nesta columna o número de formas que presenta cada un dos modelos elixidos. Malia que en teoría deberían coincidir nas cantidades o número de etiquetas e o de formas, non se dá esa correspondencia porque nós no paradigma verbal incluímos 9 formas máis, as 7 correspondentes á expresión en *-ra* do pretérito de subxuntivo e 2 máis pertencentes ás formas non normativas da primeira e segunda do plural da expresión en *-se* do pretérito de subxuntivo para as que xa existen as etiquetas.

Número de formas únicas por modelo: tomamos como referencia a cifra de 57 formas únicas por paradigma por ser a que representa os verbos regulares. Calquera dos modelos considerados que difira desta cifra será comentado no subapartado correspondente. A procedencia das 57 formas únicas tomadas como prototipo é como segue. O presente e o futuro de indicativo achegan cada un 6 formas; o copretérito, o pospretérito e mais o antepretérito de indicativo proporcionan só 5 cada un, pois neles coinciden formalmente a primeira e a terceira persoas do singular; idéntica razón adúcese para o presente de subxuntivo que achega outras 5 formas, mentres que o pretérito de subxuntivo contabiliza dúas máis (7) por incluírse no seu paradigma as formas graves non normativas da primeira e segunda persoas do plural; do pretérito de indicativo tamén extraemos 5, mais non se debe á coincidencia da primeira e terceira persoa do singular senón a que a primeira do plural coincide, excepto nos verbos que presentan temas diferenciados para o presente e o pretérito, coa súa homónima do presente de indicativo; o imperativo suma 1 soa, a da segunda plural, xa que a da segunda singular coincide graficamente coa terceira de singular do presente de indicativo; 4 achégaas o participio, 1 o infinitivo e 3 o xerundio por incluírense no paradigma deste último as dúas formas persoais; finalmente, as 4 restantes proceden do infinitivo conxugado ou se se prefire do futuro de subxuntivo e son 4 e non 5 porque a forma pertencente á primeira ou á terceira persoa do singular coincide coa do infinitivo non persoal.

Vexamos, pois, que información nos proporcionan as formas verbais:

<i>Grupos</i>	<i>Nº de lemas</i>	<i>Modelos</i>	<i>Nº de etiquetas por modelo</i>	<i>Nº de formas por modelo</i>	<i>Nº de formas únicas por modelo</i>
V1	4422	cantar	74	83	57
V2	190	comer	74	87	61
V3	223	partir	74	83	57
Vi1	1331	sacar	74	83	57
Vi2	217	cocer	74	83	57
Vi3	35	traducir	74	83	57
Vi4	30	moer	74	83	57
Vi5	6	saír	74	83	57
Vi6	144	afiar	74	83	57
Vi7	49	incluír	74	83	57
Vi8	57	advertir	74	83	57
Vi9	32	fuxir	74	86	60
Vi10	1	sorrir	74	83	51
Vi11	2	crer	74	83	57
Vi12	1	pór	74	83	62
Vi13	1	ser	74	83	63
Vi14	2	dicir	74	83	63
Vi15	1	ter	74	85	65
Vi16	1	pracer	74	83	63
Vi17	1	saber	74	83	63
Vi18	1	traer	74	83	63
Vi19	1	ver	74	83	63
Vi20	1	vir	74	89	69
Vi21	1	caber	74	83	63
Vi22	1	facer	74	83	63
Vi23	1	estar	74	83	63
Vi24	1	dar	74	83	63
Vi25	1	haber	72	82	63
Vi26	1	poñer	74	83	63
Vi27	1	ir	75	84	64
Vi28	2	poder	74	83	63
Vi29	9	lucir	74	86	60
Vi30	16	homoxeneizar	74	83	57
Vi31	6	descrer	74	83	57
Vi32	8	afacer	74	83	63
Vi33	9	abster	74	85	65
Vi34	4	prever	74	83	63
Vi35	1	reunir	74	83	57
Vi36	5	predicir	74	83	63
Vi37	12	previr	74	89	68
Vi38	29	antepor	74	83	62
Vi39	29	antepoñer	74	83	63
Vi40	1	rir	74	83	51
Vi41	11	servir	74	83	58
Vi42	5	seguir	74	86	60

Observacións sobre a columna etiquetas:

Haber non ten as 2 formas do Imperativo, de aí que nas etiquetas haxa dúas menos.

Ir presenta unha forma propia para o imperativo de primeira plural, *vamos*, polo que o número de etiquetas se incrementa nun.

Observacións sobre a columna formas:

En *comer* hai unha diferenza de catro formas porque na súa conxugación están incluídas as 4 pertencentes ao participio irregular (comesto, comesta...), mais se escolleramos o verbo *correr*, por exemplo, seguiríamos nas 83 formas por 57 únicas. Loxicamente estas 4 formas aumentan o número das formas únicas noutras tantas.

Tanto *ter* como o seu derivado *abster* presentan dúas expresións para a etiqueta Vpi20p (*tedes/abstedes* e *tendes/abstendes*), así como dúas para V0m20p (*tede/abstede* e *tende/abstende*), de aí que tanto o número de formas como o de formas únicas se incrementa en dous.

Haber non ten as 2 formas do Imperativo, non obstante presenta dúas posibilidades para a Vpi30s (*hai/ha*), motivo polo que figuran 82 formas e non 81, que serían as que lle corresponderían por non presentar formas/etiquetas para o imperativo.

Fuxir, *lucir* e *seguir* presentan 86 formas por paradigma, 3 máis do normal porque neses modelos, á parte de non coincidir a terceira singular do presente de indicativo coa segunda singular do imperativo, introducimos a maiores no presente de indicativo as formas sen alternancia vocálica para a segunda e terceira do singular, así como para a terceira do plural, co que ao lado das normativas (*foxes, foxe, foxen; loces, loce, locen; segues, segue, seguen*) aparecen as non normativas (*fluxes, fluxe, fluxen; luces, luce, lucen; sigues, sigue, siguen*). Consecuentemente, o número de formas únicas tamén se incrementa en 3 nestes modelos.

Ir incrementa nun o número de formas por posuír unha específica para a primeira persoa do plural do imperativo: *vamos*.

Vir e o seu derivado *previr* aumentan en 6 o número de formas por incluír no seu paradigma duplas expresións para as segundas de plural do presente de indicativo (*vides/prevides* e *vindes/previndes*) e do imperativo (*vide/previde* e *vinde/previnde*), así

como dobre posibilidade nas catro formas do participio (*vido, vida, vidos, vidas / previsto, prevista, previstos, previstas* e *vindo, vinda, vindos, vindas / previndo, previnda, previndos, previndas*).

Observacións sobre a columna formas únicas:

O paradigma de *rir* e mais o do seu derivado *sorrir* reducen en 6 o número de formas únicas porque neles prodúcese coincidencia gráfica entre as 5 do presente de subxuntivo e o copretérito de indicativo: *ría/sorría, rías/sorrías, riamos/sorriamos*, etc., e ademais coinciden no indicativo a primeira do singular do pretérito coa terceira do plural do presente.

Na conjugación de *servir* non se implementaron aínda as formas sen alternancia non normativas do presente de indicativo (*sirves, sirve, sirven*), polo que ao non coincidir a segunda singular do imperativo (*sirve*) con ningunha do presente de indicativo contabiliza unha forma única máis do normal e dúas menos ca os seus homónimos con alternancia *fluxir, lucir* e *seguir*.

Pór e os seus derivados, coma *antepor*, achegan 62 formas únicas por paradigma. A diferenza de 5 con respecto á maioría débese ás 5 formas máis que proporciona o futuro de subxuntivo: *puxer, puxeres, puxermos, puxerdes, puxeren*.

Os paradigmas de *ser, dicir, pracer, saber, traer, ver, caber, facer, estar, dar, haber, poñer* e os derivados *predicir, prever, afacer* e *anteponer* superan en 6 formas os modelos maioritarios con 57, debido ás 5 formas do futuro de subxuntivo diferenciadas en todos eles por formarse sobre un tema de pretérito distinto do de presente sobre o que se forma en xeral o infinitivo conxugado. Por riba, todos eles expresan no indicativo a primeira do plural do pretérito de maneira distinta á da do presente: *fomos, dixemos, prouguemos, soubemos, trouxemos*, etc.

O verbo *ir*, ademais de posuír as características explicadas no parágrafo anterior, posúe unha forma única máis por figurar no seu paradigma unha expresión propia para a primeira persoa plural do imperativo, de aí que para este paradigma se acheguen 64 formas únicas.

Ter e *abster* compórtanse coma o groso dos verbos con tema de pretérito diferenciado (63 formas), coa peculiaridade de que presentan dúas formas máis, 65 polo

tanto, debido á presenza no seu paradigma de dúas expresións distintas a maiores para a segunda de plural do presente de indicativo e do imperativo: *tendes/abstendes* e *tende/abstende*.

Finalmente *vir* (69 formas únicas) posúe unha forma única máis ca o seu derivado *previr* (68 formas únicas), porque o emprego do til diacrítico no primeiro provoca que a forma do imperativo singular (*ven*) diverxa formalmente da terceira de singular do presente de indicativo (*vén*), feito que non sucede en *previr* nin nos demais verbos. As outras 5 formas que van de 63 a 68 débense á inclusión en ambos os paradigmas de duplas expresións para as segundas de plural do presente de indicativo (*vides/prevides* e *vindes/previndes*) e do imperativo (*vide/previde* e *vinde/previnde*), así como dobre posibilidade en tres das catro formas do participio (*vido, vida, vidos, vidas / previsto, prevista, previstos, previstas, e vinda, vindos, vindas /, previnda, previndos, previndas*). As formas do participio masculino singular *vindo* e *previndo* non se erixen en formas únicas por coincidiren coas do xerundio.

2.8. Conclusións

Varias son as conclusións que podemos extraer tras a lectura deste capítulo no que describimos a estrutura interna do lexicón de XIADA, a elaboración da gramática formal do galego e a implementación das entradas léxicas pertencentes ás diferentes categorías gramaticais.

En primeiro lugar, cabe destacar a enorme complexidade que supuxo reducir a modelos formais as clases Substantivo, Adxectivo e Verbo debido á gran variedade formal que presentan estas categorías en galego. A dificultade de implementar as distintas entradas mentres non nos familiarizamos co procedemento e as etiquetas viuse compensada a longo prazo pola flexibilidade que ofrece o sistema, pois o lexicón non só serve de base para o noso analizador, senón que tamén se empregou xa para outras utilidades, como a extracción do léxico común normativo destinado a servir de alimentación para o corrector ortográfico de galego de IANUS¹²¹.

¹²¹ IANUS é a denominación que recibiu, basicamente, a ferramenta que, no marco das chamadas tecnoloxías da información e da comunicación, integra nun único sistema de información clínico-asistencial os historiais clínicos dos pacientes, para dar lugar así á chamada historia clínica electrónica. Véxase Domínguez e Mosquera Carregal (2011: 91-123).

En segundo lugar, a implementación de entradas non é un capítulo pechado. A medida que a desambiguación, manual ou automática, nos proporcione lemas non identificados irémolos incorporando paulatinamente ao lexicón, de xeito que este tamén medrará progresivamente.

En terceiro lugar, a tarefa máis complicada que levamos a cabo foi, sen ningunha dúbida, posibilitar o recoñecemento e caracterización das unidades gramaticais complexas, ou sexa, contraccións e formas verbais con enclíticos, pois á dificultade que ten de por si o tema hai que engadirlle a elevada cantidade de contraccións e de combinacións de clíticos que son posibles en galego.

Finalmente, aínda sentíndonos satisfeitos polos resultados obtidos, non podemos dar por pechado este capítulo sen facer fincapé na necesidade de seguir elaborando regras lingüísticas que melloren o comportamento do analizador e etiquetador nin esquecérmonos de que tamén está pendente a implementación masiva de unidades multipalabra variables.

Capítulo 3

A etiquetaxe automática

3.1. *Introdución*

Desde o punto de vista lingüístico, unha vez que se establece o etiquetario e se constrúe o lexicón computacional implementándoo cun número aceptable de entradas clasificadas morfoloxicamente segundo os parámetros descritos no *tagset*, xa se está en disposición de etiquetar automaticamente calquera texto galego contemporáneo. Trátase, simplemente, de executar o programa axeitado: o etiquetador.

Computacionalmente, antes de proceder a etiquetar un documento, é necesario primeiro delimitar as distintas unidades léxicas que este contén, tarefa que como xa vimos no capítulo anterior lle corresponde ao preprocesador. Este é un labor que, aínda que poida parecer sinxelo, non resulta trivial, e moito menos para o galego, pois hai que indicarlle a un programa como segmentar formas verbais con pronomes enclíticos ou como separar os formantes dunha contracción. Seguimos traballando para que cometa menos erros, pero é inevitable que nalgúns casos de ambigüidade segmental complexa falle de vez en cando. É, sobre todo, un traballo informático no que os lingüistas participan coa construción de regras que gobernan o comportamento do preprocesador. Por exemplo, hai unha década, cando se estaba comprobando o funcionamento das formas verbais con clíticos, atopámonos con que para a unidade gráfica *semana* o sistema XIADA propoñía unha análise na que segmentaba o *token* en “se” raíz verbal, “ma” contracción dos pronomes *me* mais *a*, e o último constituínte formábao o alomorfo do pronome acusativo de terceira persoa “na”. Vimos entón a necesidade de impedir análises coma esta, polo que creamos unha regra na que se indica que na mesma secuencia non pode haber dous clíticos de acusativo e, ademais, para *no*, *na*, *nos*, *nas* –pronome átono acusativo de terceira– só será unha etiqueta posible se vai enclítico a unha forma verbal rematada en ditongo, non en ningunha outra circunstancia.

No caso de XIADA, así mesmo, o preprocesador, tras delimitar as unidades léxicas presentes nos textos, asígnalle a cada unha os lemas e etiquetas posibles. Deste

xeito a saída do documento preprocesado achéganos o texto analizado contendo para cada unidade todas as posibles análises morfolóxicas. Por exemplo, o preprocesador ante unha unidade ortográfica coma *polos*, na que non só hai ambigüidade morfolóxica –máis dunha etiqueta–, senón que tamén existe nela ambigüidade segmental –varía a segmentación da unidade en constituíntes–, proporciona as seguintes análises:

- Unha análise-alternativa na que se considera o *token* unha única unidade gramatical e se clasifica esta como substantivo común masculino plural cuxo lema é *polo*:

```
<alternativa>
  <constituínte>
    <forma>polos</forma>
    <etq_ lema>
    <etiqueta>Scmp</etiqueta>
    <lema>polo</lema>
    </etq_ lema>
  </constituínte>
</alternativa>
```

- Outra alternativa na que se segmenta a unidade ortográfica en dous constituíntes, caracterizando o primeiro como a preposición *por* e o segundo como o artigo determinado masculino plural cuxo lema é *o*:

```
<alternativa>
  <constituínte>
    <forma>por</forma>
    <etq_ lema>
    <etiqueta>P</etiqueta>
    <lema>por</lema>
    </etq_ lema>
  </constituínte>
  <constituínte>
    <forma>los</forma>
    <etq_ lema>
    <etiqueta>Ddmp</etiqueta>
    <lema>o</lema>
    </etq_ lema>
  </constituínte>
</alternativa>
```

- Finalmente, unha última alternativa na que se desagrega a unidade gráfica *polos* no elemento gramatical *pos*, analizado como segunda persoa do presente de indicativo dos verbos *pór* e *poñer*, e mais no elemento gramatical *los*, caracterizado como forma do pronome átono de acusativo de terceira persoa masculino plural cuxo lema é *o*:


```

<alternativa>
  <constituínte>
    <forma>pos</forma>
    <etq_ lema>
    <etiqueta>Vpi20s</etiqueta>
    <lema>poñer</lema>
    </etq_ lema>
    <etq_ lema>
    <etiqueta>Vpi20s</etiqueta>
    <lema>pór</lema>
    </etq_ lema>
  </constituínte>
  <constituínte>
    <forma>los</forma>
    <etq_ lema>
    <etiqueta>Raa3mp</etiqueta>
    <lema>o</lema>
    </etq_ lema>
  </constituínte>
</alternativa>

```

Chegado este punto teríamos o documento analizado con todas as etiquetas e lemas posibles mais, dado que na etiquetaxe de corpus non interesan todas as análises potenciais dunha unidade, senón só a análise que corresponde en cada caso concreto, precisamos dun etiquetador que escolla automaticamente a etiqueta e o lema correctos. Como conseguilo vai centrar os nosos esforzos neste capítulo. Así, veremos cales son as principais características da ferramenta que empregamos para etiquetar e daremos conta dos factores que inciden na súa taxa de acerto: tamaño do etiquetario, robustez do corpus de adestramento, tratamento das formas descoñecidas e distintos tipos de ambigüidade cos que se ten que enfrontar.

3.2. O Etiquetador

3.2.1. Características

A ferramenta que empregamos para a etiquetaxe morfosintáctica de corpus é a denominada *Etiquetador/Lematizador do Galego Actual*, XIADA¹²², un etiquetador estatístico baseado nos modelos de Markov de grao 2 (Markov, 1913: 153-162) desenvolvido conxuntamente polo Centro Ramón Piñeiro para a Investigación en Humanidades e o grupo COLE das universidades da Coruña e Vigo e destinado a etiquetar os documentos do *Corpus de Referencia do Galego Actual* co obxecto de

¹²² <http://corpus.cirp.es/xiada>

proporcionar recursos e ferramentas para a análise lingüística computacional do galego actual.

A etiquetación debe entenderse como o proceso no que as distintas unidades léxicas dun documento reciben en primeira instancia todas as análises posibles e logo, nunha segunda fase, consecutiva no tempo mais integrada no sistema e polo tanto non separable da anterior, a elección da análise morfosintáctica que se considera que é a correcta en cada ocorrencia; é dicir, a etiquetaxe de corpus implica a anotación –asignación de todas e cada unha das etiquetas correspondentes a unha unidade considerada como elemento illado– e a desambiguación –escolla dunha única etiqueta e lema de entre todas as posibilidades para unha unidade considerada no seu contexto–.

Para acometer a etiquetación pódese partir dunha aproximación lingüística ou dunha estatística. Segundo Voutilainen (1999: 218-219), é preferible partir da aproximación lingüística, posto que nunha de tipo probabilístico precísase un corpus de adestramento coidadosamente etiquetado de, polo menos, decenas de miles de palabras, aínda que para ben ser se necesitarían centos de miles. A maiores, segundo este autor, a opción estatística presenta outras dúas limitacións: por unha banda, o seu teito de acerto sitúase no 97% e, pola outra, se o estilo dos documentos a etiquetar non está representado no corpus de adestramento o acerto do etiquetador diminuirá de maneira considerable.

Agora ben, á parte de que este último factor tamén se lle pode reprochar aos etiquetadores baseados en modelos lingüísticos, tampouco a aproximación lingüística é unha panacea, pois, malia que Voutilainen (1999: 217-246) se decante por ela, entre as súas desvantaxes están, tal e como manifesta Civit, as seguintes:

- (i) el elevado coste de la construcción del modelo de reglas; (ii) la poca transportabilidad de una lengua a otras; y (iii) el hecho de que puede ser poco robusto si el modelo que se ha construído no es lo suficientemente completo. [Civit, 2003: 136]

Graña (2000) describe os distintos sistemas utilizados para a etiquetación de textos en linguaxe natural e avalía catro deles: dous que empregan como principio de funcionamento os modelos de Markov ocultos (TNT e GALENA), un sistema baseado en regras (BRILL) e finalmente un sistema híbrido que utiliza fundamentos das dúas aproximacións anteriores (JMX). Na súa análise comparativa-contrastiva destácase a

preferencia tradicional das aproximacións puramente estocásticas fronte ás aproximacións baseadas en regras, debido a dous factores primordiais: o seu bo rendemento e, sobre todo, as facilidades de adestramento automático. Remata este autor o seu estudo con dúas conclusións principais sobre o proceso de etiquetación de textos en linguaxe natural:

- resulta máis adecuado o marco probabilístico ca as aproximacións simbólicas;
- os sistemas de etiquetación puramente estocásticos deben empregarse en combinación con técnicas de análise sintáctica robusta que reduzan a pequena porcentaxe de palabras etiquetadas erroneamente que sempre están presentes nos sistemas probabilísticos –entre o 2% e o 3%–.

Así pois, a opción pola que nos decantamos para XIADA é a dun etiquetador automático estatístico que, completado con regras lingüísticas que o axuden a resolver as ambigüidades segmentais e morfolóxicas que non logra solucionar o método baseado na estatística, pretende superar o rango máximo de rendemento ao que pode aspirar un sistema deste tipo: 95%-97% de acerto.

A continuación analizaremos os diferentes factores que inciden na taxa de acerto dun etiquetador.

3.2.2. Rendemento e precisión

As cifras de rendemento coñecidas para os etiquetadores atópanse situadas, en xeral, dentro do rango do 95 ao 97% de acerto (Graña, 2000: 137-165).

Agora ben, o rendemento e precisión dun etiquetador non depende só da valía da ferramenta en si mesma e do método escollido para etiquetar, senón que existen outros factores, externos á propia ferramenta, que van condicionar considerablemente os resultados. Entre eles destacan os seguintes (Graña, 2000: 6-7):

- O xogo de etiquetas ou etiquetario. Normalmente canto máis grande é o conxunto de etiquetas considerado, máis ambigüidade potencial existe. Ademais, a tarefa da etiquetaxe vólvese máis complexa e aumenta a probabilidade de erro calculado sobre o conxunto de todas as palabras dun texto.

- A cantidade de texto de adestramento dispoñible. En xeral, de canto máis texto se dispoña, mellor é o resultado.
- A diferenza entre, por un lado, o corpus de adestramento e o dicionario ou lexicón e, por outro, o corpus de aplicación. Se os textos de adestramento e os textos que posteriormente se van etiquetar proceden da mesma fonte ou mostran un estilo semellante, a precisión será alta. Pola contra, se os textos de aplicación pertencen a un período distinto ou a un xénero ou estilo diferentes, por exemplo textos xornalísticos contra textos científicos, o rendemento será baixo.
- As palabras descoñecidas. Depende en gran medida da amplitude do lexicón e do tipo de texto que se pretenda etiquetar.

Deseguido iremos desmiuzando cada un destes puntos e vendo como responde a eles o sistema XIADA.

3.2.2.1. Tamaño do etiquetario

Polo que respecta ao tamaño do etiquetario, xa vimos no Capítulo 1 que dende EAGLES se propugna non superar as 100 etiquetas:

Experience (Church, UPenn Treebank, IBM France, etc.) shows that the tagset should be under 100 tags. In fact, the Penn Treebank project collapsed many tags compared to the original Brown tagset, and got better results. [Monachini e Calzolari, 1996]

Así mesmo, segundo Jurafsky (2000: 296-297), os etiquetarios máis usados para o inglés son o *Penn Treebank*¹²³ que consta de 45 etiquetas e foi empregado para a etiquetaxe do corpus *Brown*¹²⁴, o *C5* de 61 etiquetas empregado polo programa *CLAWS*¹²⁵ para etiquetar o *British National Corpus* (BNC)¹²⁶ e o *C7*, de 146 etiquetas, utilizado para etiquetar un subconxunto do BNC.

Seguindo esta mesma tónica de simplificación, un exemplo drástico ten lugar no etiquetario utilizado polo analizador estatístico para a lingua galega desenvolvido no Departamento de Teoría de la Señal y Comunicaciones da Universidade de Vigo,

¹²³ Véxase <http://www.cis.upenn.edu/~treebank>

¹²⁴

http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html

¹²⁵ <http://ucrel.lancs.ac.uk/claws/>

¹²⁶ <http://www.natcorp.ox.ac.uk>

empregado na etiquetaxe morfosintáctica dun corpus de galego destinado a aplicacións en sistemas de recoñecemento e síntese de voz (Méndez Pazó et al., 2003a e 2003b; Seijo Pereiro et al., 2004) onde se produce unha diminución do 93,03%, xa que das 760 etiquetas dispoñibles para a anotación pasan a dispoñer tan só de 53 na etiquetaxe, coas que se acada, quedémonos con este dato, o 97% de acerto no proceso de desambiguación.

Non obstante, non todos os sistemas de etiquetación seguen as pautas anteriores; pola contra, existen outros que presentan un etiquetario equiparable en tamaño ao que describimos no Capítulo 1, ou incluso superior. Así, o *tagset* da versión española do corpus ITU (*International Telecommunications Union CCITT Handbook*), desenvolvido no marco do proxecto CRATER¹²⁷, consta de 492 etiquetas (Sánchez León, 1994); tamén para o castelán, o etiquetario do sistema GALENA (Graña, 2000: 233-235) emprega 373 etiquetas diferentes e, xa para a nosa lingua, á parte das 383 de XIADA, o sistema do CTAG utiliza un xogo de 225, mentres que o do TECTRA, sen ter en conta as que resultan da combinación de etiquetas nas distintas amálgamas, revela 254 simples.

Un punto e unha décima separan a porcentaxe de acerto na etiquetaxe de corpus en galego con 53 etiquetas (97%, o aplicado en recoñecemento e síntese de voz da Universidade de Vigo) e con 383 (95,99%, o de XIADA aplicado no CORGAetq¹²⁸). Así pois, se ben é certo que o tamaño do etiquetario dificulta as posibilidades de acerto na etiquetaxe, non parece que por si só sexa un factor determinante e provocador dunha alta porcentaxe de erro. É máis, deixando de lado a finalidade da etiquetaxe dos dous corpus cuxa taxa de acerto acabamos de comparar en relación co tamaño de cadanseu etiquetario, un punto e unha décima de diferenza compensa manter etiquetarios desiguais para anotación e para etiquetaxe de corpus?, compensa realmente reducir o etiquetario? Cremos que non, cremos que o incremento do acerto hai que buscalo noutra dirección: as regras lingüísticas deben asumir ese reto.

¹²⁷ <http://www.comp.lancs.ac.uk/linguistics/crater/corpus.html>

¹²⁸ Domínguez et al., 2008.

3.2.2.2. O corpus de adestramento

En primeiro lugar, no referente ao corpus de adestramento, sabemos que para poder empregar con certas garantías un etiquetador automático estatístico necesitamos dispoñer dun conxunto de textos etiquetados, canto máis amplo e diversificado lingüisticamente, mellor, para que o sistema conte con datos nos que apoiarse para realizar as súas eleccións.

Así mesmo, é fácil deducir que as probabilidades de acerto do etiquetador aumentarán proporcionalmente en relación coa amplitude e mais a variedade temática e estilística do corpus etiquetado que se empregue para exercitalo. Non obstante, antes de seleccionar os documentos que van servir de base para instruír o etiquetador, existe un factor fuxidío que non debemos perder de vista. Trátase da necesidade de que todas as etiquetas que aparecen no lexicón ou susceptibles de apareceren nun texto estean documentadas no corpus de adestramento.

En xeral, aínda que os etiquetadores poden ser aplicables a calquera conxunto de etiquetas, na maioría deles non se pode especificar de maneira explícita ese conxunto. Neses casos, o etiquetario dedúcese automaticamente o etiquetador a partir das distintas etiquetas que figuran no corpus de adestramento. Os problemas xorden cando se utiliza un dicionario externo que presenta algún par unidade-etiqueta que non consta no corpus de adestramento, pois nestas ocasións o etiquetador atópase cunha etiqueta descoñecida para a que non posúe ningunha información.

O sistema de XIADA, a semellanza do de GALENA (Graña, 2000), permite explicitar previamente o xogo de etiquetas, comproba que todas as etiquetas presentes pertencen a ese conxunto e esixe que cada unha das etiquetas apareza no corpus de adestramento polo menos unha vez.

O máis usual é que o conxunto de documentos co que se constrúe un corpus de adestramento se escolla aleatoriamente, mais este sistema de selección non garantiza a aparición no corpus de todas e cada unha das etiquetas. Por este motivo o corpus de adestramento do sistema XIADA presenta dúas orixes e motivacións distintas:

- 1) Por unha banda, para acadar o obxectivo de que no corpus de adestramento aparezan todas as etiquetas, creamos un subcorpus artificial cuxa característica

principal é a de conter como mínimo unha ocorrencia de cada unha das etiquetas. A este subconxunto de texto etiquetado semiautomáticamente denominámolo *kernel* ou *núcleo* de XIADA.

- 2) Por outra banda, para lograr a representatividade dos diferentes xéneros e estilos presentes no CORGA, corpus sobre o que imos aplicar o etiquetador automático, extraemos deste un subcorpus de entre 150 000 e 200 000 formas gráficas de cada un dos grandes bloques do CORGA: prensa, ficción e ensaio.

O corpus de adestramento de XIADA non está aínda concluído. Polo momento, construíuse o que os especialistas denominan TC0, o *Training Corpus Zero* ou Corpus de Adestramento Cero, que nós chamamos *núcleo* ou *kernel* do corpus de adestramento. Finalizouse tamén co bloque correspondente ao estilo xornalístico e, nos próximos anos, o obxectivo máis inmediato é cubrir o apartado relativo á ficción. Unha vez que se remate con este último, procederase a etiquetar automaticamente pequenas porcións de textos ensaísticos, analizando con detalle os resultados, para determinar se é necesario incrementar o corpus de adestramento con outros estilos e rexistros.

Vexamos agora cales son as características concretas de cada un dos módulos que conforman, polo momento, o corpus de adestramento de XIADA.

3.2.2.2.1. O núcleo de Xiada

O *kernel* ou núcleo de XIADA está constituído por un arquivo en formato texto, *nucleo_etiquetado.txt*, no que se recolle a etiquetaxe de 265 oracións nas que conseguimos reunir a aparición de cada unha das distintas etiquetas polo menos unha vez. O arquivo ocupa 75 766 caracteres e dispón de 5788 liñas, das cales 265, constituídas só por unha marca de parágrafo, sinalan simplemente a separación entre cada unha das oracións e as 5523 restantes responden á análise das distintas unidades gramaticais do documento; en concreto, 2158 delas corresponden a análises diferentes. A sintaxe de cada liña é como segue:

unidade gramatical tabulador etiqueta tabulador lema

As 383 combinacións correctas dende un punto de vista lingüístico que se xeneran a partir do etiquetario implican neste subcorpus a ocorrencia de 1921 formas ou unidades gramaticais diferentes remisibles a 1518 lemas distintos.

As características do dicionario constituído polas 1921 formas distintas que aparecen no Núcleo de XIADA son as seguintes:

Total de formas distintas:	1921
Total de formas+etiquetas+lemas diferentes:	2143
Formas con 1 etiqueta:	1776
Formas con 2 etiquetas:	109
Formas con 3 etiquetas:	16
Formas con 4 etiquetas:	7
Formas con 5 etiquetas:	5
Formas con 6 etiquetas:	4
Formas con 9 etiquetas:	2
Formas con 13 etiquetas:	1

Isto é, seguindo a Graña (2000: 140), 1921 formas distintas, con 2143¹²⁹ etiquetas posibles. Se con estes datos calculamos a porcentaxe de formas ambiguas e o número medio de etiquetas por forma, obtemos:

$$\% \text{ formas ambiguas} = \frac{\# \text{ formas ambiguas}}{\# \text{ formas}} \times 100 = \frac{1921 - 1776}{1921} \times 100 = 7,54\%$$

$$\# \text{ nº medio de etiquetas por forma} = \frac{\# \text{ etiquetas}}{\# \text{ formas}} = \frac{2143}{1921} = 1,11 \text{ etiquetas por forma}$$

Na táboa seguinte recollemos as unidades léxicas do núcleo de XIADA que se correlacionan con máis dunha etiqueta. A primeira columna de cada serie acolle o número de etiquetas que recibe neste subcorpus (Nº) e a segunda a forma ortográfica correspondente (Forma)¹³⁰:

¹²⁹ Esta cifra resulta de multiplicar o número de formas polo das etiquetas distintas que leva cada unha delas.

¹³⁰ A ordenación é a natural nun procesador de textos: primeiramente os numerais, logo as unidades con maiúscula inicial e, finalmente, as minúsculas.

Nº	Forma
2	17
2	20
2	25
2	Estado
2	Marcial
2	PSOE
2	Segura
2	UXT
2	Vde.
2	acordo
2	algunha
2	algún
2	as
2	ben
2	cabeza
2	canta
2	certo
2	compra
2	común
2	condena
2	cámaras
2	denuncia
2	difíciles
2	directamente
2	dobre
2	dous
2	dúas
2	el
2	ela
2	elas
2	eles
2	elo
2	esas
2	españóis
2	especialista
2	esta
2	estas
2	este
2	estes
2	estudiantes
2	faga
2	fixera
2	fun
2	galegos
2	h.
2	importante
2	la
2	las
2	lo
2	locais

Nº	Forma
2	los
2	maiores
2	mais
2	mañá
2	medio
2	meu
2	meus
2	mira
2	miña
2	miñas
2	moitas
2	moitos
2	máis
2	ningunha
2	ninguén
2	ningún
2	non
2	nosa
2	nosas
2	noso
2	nosos
2	nosoutros
2	o
2	os
2	outra
2	outras
2	periodistas
2	pois
2	políticos
2	poña
2	primeira
2	primeiras
2	primeiros
2	quén
2	se
2	segundo
2	si
2	socialistas
2	son
2	tal
2	tan
2	terceiro
2	teu
2	teus
2	todas
2	todo
2	todos
2	técnico
2	túa
2	túas

Nº	Forma
2	vai
2	veu
2	vosa
2	vosas
2	voso
2	vosos
2	vosoutras
2	vosoutros
2	vostedes
3	a
3	auxiliar
3	cada
3	che
3	como
3	eu
3	lles
3	mal
3	mesmo
3	min
3	moito
3	outro
3	outros
3	qué
3	trece
3	vostede
4	calquera
4	catro
4	mellor
4	tres
4	un
4	unha
4	vós
5	lle
5	nós
5	quen
5	súas
5	tí
6	seu
6	seus
6	súa
6	te
9	me
9	nos
9	vos
13	que

As secuencias presentes no núcleo etiquetado do sistema XIADA hai que procuralas en catro fontes distintas. Quixemos, na medida do posible, exemplificar todas as etiquetas do noso etiquetario con casos reais tirados do corpus que pretendemos etiquetar, o CORGA; mais en última instancia, cando a pescuda obrigaba a revisar centos de resultados para confirmar un, recorreremos a outras fontes. Ordenadas polo seu peso no subcorpus nuclear final, son as seguintes:

1) O CORGA. As etiquetas máis comúns ou as facilmente localizables por caracterizar algunha forma ortográfica moi definida foron documentadas a partir de oracións extraídas do CORGA. Por exemplo, os nomes propios de persoa *Mercedes*, *Jesús* e *Marcial* representan as etiquetas correspondentes a Substantivo propio feminino e masculino, respectivamente, cuxo número non aplica (Spf0 e Spm0).

Temeroso de que poideran contar algo, El Pulpo acabou con Mercedes e Jesús, e para isto intimidou a Marcial. [CORGA: CG1999-05-13/32]

Outra mostra. A ocorrencia dun indefinido nuclear, invariable na súa forma coma *alguén* ou *ninguén*, con valor plural atópase con bastante facilidade no CORGA coa consulta á versión en liña poñendo na caixiña das *Palabras* “ninguén *mos” e en *Tipo de busca* “Frase exacta”:

Ninguén sospeitabamos que as medidas a tomar foran xa as tan mortuorias dos nichos de Bonaval. [CORGA: CG1999-05-21/7]

Un par de exemplos máis desta fonte. O pronome tónico en caso *nominativo* correspondente á terceira persoa e xénero neutro ([Rtn3ns] *Pronome, tónico, nominativo, terceira, neutro, singular*), *elo*, localízase dun xeito sinxelo coa consulta ao CORGA na caixiña das *Palabras* “que elo” e en *Tipo de busca* “Frase exacta”. Dos casos que cumprían o requisito da busca, ao chou, foi seleccionado o último deles:

Tamén é un medio bo para combatilo, fomentar as nosas pequenas ambicións, loitar para realizalas cabalmente e amar o esforzo que elo require. [CORGA: CG2002-05-14/29]

A sorte tamén xogou unha baza importante na selección de secuencias que contiveran algunha etiqueta determinada. Así, por exemplo, a aparición da oración na que se documenta a primeira persoa do singular do pospretérito de indicativo debeuse á busca de erros que se nos puideran escapar na construción do CORGA producidos pola mala lectura do recoñecedor óptico de caracteres (OCR) utilizado para a dixitalización

de documentos. Periodicamente levamos a cabo unha revisión do corpus consultando nel unha listaxe de erros sistemáticos entre os que se atopan a lectura de *escoltar* por *escoitar*. De aí procede a detección de

Se viñesen a min escoitáriaos”, dixo. [CORGA: CG1999-05-04/21]

2) Creación propia. Cando a localización dalgunha das etiquetas resultaba moi dificultosa, recorremos á invención do exemplo. Así, a aparición dun substantivo masculino ou feminino invariable en número, por poñer un caso, non é tan inusual; mais a combinación dunha unidade deste tipo cun adxectivo común en canto ao xénero e invariable en número si é moi infrecuente. Ante as consultas infrutuosas realizadas en distintas fontes, optamos por inventar dúas secuencias, inverosímiles pero posibles, nas que comparecesen as etiquetas antes mencionadas (*Scfa* seguida de *A0fa* e *Scma* seguida por *A0ma*):

¡Sífilis mexericas!

¡Parabrisas furafollas!

Non tan absurda pero si igual de ilóxica é a oración coa que documentamos a concorrencia de seis das etiquetas existentes para posesivos (*Md2smp*, *teus*; *Md2sfp*, *túas*; *Md2pmp*, *vosos*; *Md2pfp*, *vosas*; *Md2pms*, *vosos*; *Md1sfp*, *miñas*):

Cos teus cans, as túas gatas, os vosos cabalos, as vosas vacas, o voso elefante e as miñas galiñas xa teriades unha boa granxa se quixesedes montala.

Noutras ocasións as secuencias creadas resultan do máis verosímil e corrente na nosa lingua. A dificultade nos seguintes exemplos residía na localización dun pronome átono de primeira e segunda plural cuxo valor casuístico fose *formante léxico* e o seu referente fose un feminino:

Atrevémonos a facelo nó-las dúas?

Neniñas, non debes espreguizarvos diante de todos!

3) A *Gramática Galega* (Álvarez, Regueira e Monteagudo, 1986: 382). Dada a dificultade de encontrar ocorrencias para certas etiquetas, como son as pertencentes ao futuro de subxuntivo, tomamos como modelo ou copiamos literalmente algúns dos exemplos desta gramática. As seguintes oracións exemplifican esta fonte:

Xuro que se non morrer aquí hei volver.

Onde vires moito fume non te vaias quentar.

Como sementarmos, así colleremos cando seguemos.

4) *Gramática da lingua galega* (Álvarez e Xove, 2002: 319). Se a memoria non nos traizo, só botamos man desta obra para documentar a segunda persoa do plural do denominado xerundio conxugado:

Creo que todo sairá mellor facéndodes como di el.

Dada a artificialidade e as diversas fontes do núcleo ou *kernel* de XIADA, loxicamente, este non forma parte do *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq), mais si está dispoñible para a súa descarga na páxina correspondente ás descargas do proxecto XIADA¹³¹.

A versión descargable contén, ademais do arquivo *nucleo_etiquetado.txt* que xa describimos no comezo desta epígrafe, outro arquivo denominado *nucleo_oracions.txt*, no que se recollen, sen codificación nin análise de ningún tipo, co formato de unha por liña, as 265 oracións que, etiquetadas, constitúen o corazón de XIADA. O tamaño deste arquivo é de 26 795 caracteres e consta de 4476 palabras, é dicir, unha media de 16 palabras por oración.

3.2.2.2.2. O subcorpus xornalístico de ámbito temático específico

O subcorpus escollido para o adestramento inicial foi extraído do CORGA e constou de 200 000 palabras procedentes de noticias xornalísticas do ámbito económico, pertencentes tanto a revistas, as menos (18), como a xornais, as máis (573). O período temporal nel recollido abrangue dende o ano 1995 ata o 2002.

A especificidade temática deste bloque débese á nosa participación no subproxecto de investigación coñecido como GARI-COTER¹³², englobado á súa vez no proxecto RICOTERM¹³³, en cuxos obxectivos se incluía a constitución dun corpus textual de economía en galego, codificado no estándar XML e procesado lingüisticamente. Dado que a reutilización de recursos era unha baza importante nos fundamentos do proxecto e que no Centro Ramón Piñeiro para a Investigación en

¹³¹ <http://corpus.cirp.es/xiada/descargas.html>

¹³² Creación e integración multilingüe de recursos lingüísticos en galego para Recuperación de Información mediante estratexias de control terminolóxico e discursivo en ámbitos comunicativos especializados, Ministerio de Educación e Ciencia (HUM2004-05658-C02-02), 2004-2007.

¹³³ Véxase http://ricoterm.iula.upf.edu/2/index_gal.htm [consultado o 28/12/2010]

Humanidades (CIRP) se dispoñía non só das ferramentas que permitirían procesalo senón tamén dun corpus lingüístico, o CORGA, que incluía cantidade suficiente de textos do ámbito económico, tanto xenéricos como especializados, decidiuse construír dous corpus: un corpus de prensa económica e un corpus especializado de economía. Ambos foron etiquetados automaticamente co *Etiquetador/Lematizador do Galego Actual*, mais só o primeiro gozou dunha revisión manual desa etiquetaxe.

A nosa pretensión era aproveitar a etiquetaxe do subcorpus xenérico de prensa de GARI-COTER para constituír o comezo do que debería ser o corpus de adestramento do sistema de XIADA, non obstante, causas alleas ao que aquí tratamos impedíronnos lograr este desiderátum.

Porén, máis tarde decidimos de todas maneiras levar adiante a revisión manual da etiquetación efectuada dun xeito automático sobre o conxunto de documentos que conforman o subcorpus económico xenérico de prensa, pois a extracción aleatoria estaba feita e os documentos, etiquetados. Consideramos ademais que era unha boa oportunidade, dado o estilo e temática dos textos, de comprobar como se comportaba o etiquetador e confiamos dende o primeiro momento en que o microcorpus desambiguado fose útil para a aprendizaxe das estruturas sintácticas características do xénero xornalístico.

Este subcorpus está constituído por 601 arquivos en formato XML correspondentes a outras tantas noticias pertencentes ás seguintes publicacións:

A Nosa Terra: 22 noticias

O Correo Galego: 551 noticias

Feiraco: 6 noticias

Galicia Internacional: 8 noticias

Man Común. Revista galega mensual de información xeral: 1 noticia

Petroglifo. Revista da Asociación de Empresarios de Artes Gráficas de Galicia: 2 noticias

Teima. Revista Galega de Información Xeral: 1 noticia

Tempos Novos: 10 noticias

Non imos describir aquí o proceso formal de desambiguación manual realizado en XML, pois esa información non é pertinente para o aspecto que estamos tratando; polo que respecta ás decisións lingüísticas tomadas en casos ambiguos ou problemáticos, verémolas no capítulo seguinte.

Lembremos, ademais, que o corpus de adestramento do sistema XIADA, fóra o núcleo, e mais o CORGAetq coinciden nos documentos analizados, mais non así nas utilidades dun e doutro.

No CORGAetq, coma en calquera corpus, ao realizar unha consulta recuperamos o enunciado, máis ou menos extenso, no que o obxecto da busca aparece recollido. Recuperamos, así mesmo, a unidade gráfica na que está inmerso o elemento gramatical analizado e toda a información dispoñible relativa á localización física desa ocorrencia: autor, obra, ano, etc.

Pola contra, no corpus de adestramento, sexa o núcleo, sexa o xornalístico económico ou sexa o xenérico de prensa, a única información coa que contamos é a de unidade gramatical e etiqueta coa que esta se asocia, posto que os modelos estatísticos de Markov están pensados soamente para procesar os *tokens* coas súas etiquetas e nada máis. O etiquetador, daquela, obvia as unidades e tampouco usa a información relativa aos lemas; estes asóciense coas etiquetas pero o etiquetador, ante unha forma coa mesma etiqueta susceptible de remitirse a máis dun lema, colle un ao azar. O subcorpus que nesta epígrafe imos describir está formado, en consecuencia, só por unidades gramaticais coas correspondentes etiquetas e lemas.

A semellanza do núcleo de XIADA, o subcorpus de prensa económica está dispoñible para a súa descarga en dous arquivos, un coas oracións (*corpus_economia_prensa_oracions.txt*) e outro coas análises desas oracións (*corpus_economia_prensa_etiquetado.txt*)¹³⁴. A maiores, inclúese un arquivo, *corpus_economia_prensa_etiquetado.xml*, en formato XML no que tamén se recolle en cada caso a unidade obxecto de análise. A descrición que nós realizamos efectúase sobre o arquivo *corpus_economia_prensa_etiquetado.txt* e os exemplos que recollemos proceden da versión 2.4 do CORGAetq.

¹³⁴ <http://corpus.cirp.es/xiada/descargas.html>

O arquivo que serve para adestrar o etiquetador con este subcorpus ocupa 3 384 023 caracteres e consta de 242 964 liñas, das cales 7 633 se corresponden con marcas de parágrafo que sinalan o remate da análise dunha oración e o comezo da seguinte. Os elementos gramaticais analizados son, polo tanto, 235 331, dos cales 21 981 constitúen unidades de análise distintas e implican a ocorrencia de 19 868 formas ou unidades gramaticais diferentes, remisibles a 13 542 lemas distintos.

A sintaxe de cada liña segue sendo

unidade gramatical tabulador etiqueta tabulador lema

Ademais, só 252 etiquetas das 383 ás que dá orixe o noso etiquetario están representadas neste subcorpus, o que significa que 131 non se empregan neste bloque. Por unha banda, isto demostranos que a tese sobre a que fundamentamos a construción do núcleo de XIADA é atinada: a selección aleatoria dun conxunto de textos, por moi amplo que este sexa, non garante que apareza documentado por completo o xogo de etiquetas determinado para unha lingua dada. Por outra banda, non é preciso recorrer a etiquetas infrecuentes coma as que clasifican os usos do futuro de subxuntivo; outras tan comúns coma as que catalogan o posesivo de primeira ou segunda persoa singular en masculino plural (Md1smp, Mn1smp; Md2smp, Mn2smp) non aparecen nin unha soa vez neste subcorpus de algo máis de 200 000 palabras; o mesmo sucede coas formas verbais pertencentes á segunda persoa, ben do singular ben do plural, do infinitivo conxugado, copretérito, pospretérito e antepretérito de indicativo ou pretérito de subxuntivo.

A distribución de formas por número de etiquetas no módulo de economía é como segue:

Total de formas distintas:	19 868
Total de formas+etiquetas+lemas diferentes:	21 980
Formas con 1 etiqueta:	18 203
Formas con 2 etiquetas:	1363
Formas con 3 etiquetas:	209

Formas con 4 etiquetas:	64
Formas con 5 etiquetas:	20
Formas con 6 etiquetas:	4
Formas con 6 etiquetas:	4
Formas con 7 etiquetas:	1
Formas con 8 etiquetas:	2
Formas con 9 etiquetas:	1
Formas con 12 etiquetas:	1

Se cos datos obtidos, seguindo a Graña (2000: 140), calculamos a porcentaxe de formas ambiguas e o número medio de etiquetas por forma, obtemos as seguintes cifras:

$$\% \text{ formas ambiguas} = \frac{\# \text{ formas ambiguas}}{\# \text{ formas}} \times 100 = \frac{19\ 868 - 18\ 203}{19\ 868} \times 100 = 8,38\%$$

$$\# \text{ nº medio de etiquetas por forma} = \frac{\# \text{ etiquetas}}{\# \text{ formas}} = \frac{21\ 980}{19\ 868} = 1,10 \text{ etiquetas por forma}$$

No cadro seguinte recolleemos as unidades léxicas do subcorpus xornalístico do ámbito económico que se correlacionan con tres ou máis etiquetas. A primeira columna de cada serie acolle o número de etiquetas que se lle asignan (Nº), mentres que na segunda indícase a forma ortográfica correspondente (Forma):

Nº	Forma
3	1
3	1,1
3	1,5
3	1,50
3	1,8
3	1,9
3	1.000
3	110
3	111
3	119
3	12,4
3	12,5
3	120.000
3	124

Nº	Forma
3	13
3	14
3	150
3	16
3	17
3	18
3	180
3	2,2
3	2,6
3	2,8
3	2.000
3	20.000
3	23
3	24

Nº	Forma
3	250
3	250.000
3	27
3	3.000
3	30.000
3	33
3	34
3	36
3	41
3	43
3	44
3	5.000
3	50.000
3	500.000

Nº	Forma
3	54
3	59
3	6
3	62
3	7
3	7,1
3	7.500
3	70
3	72
3	75.000
3	8
3	800
3	84
3	86
3	90
3	93
3	95
3	97
3	98
3	CCAA
3	CEG
3	Fidac
3	a
3	adiante
3	alerta
3	anual
3	auxiliares
3	axentes
3	baixa
3	baixo
3	capaz
3	capital
3	cara
3	central
3	cero
3	certo
3	claro
3	como
3	común
3	conforme
3	conscientes
3	continental
3	corrente
3	correspondentes
3	crecente
3	cuarta
3	cuarto
3	cultural
3	cómo
3	dependente
3	determinante
3	diferencial
3	diferente
3	diferentes

Nº	Forma
3	difícil
3	dobre
3	doce
3	dous
3	empregada
3	empresariais
3	encargada
3	encargado
3	entrevistados
3	españóis
3	esta
3	estadounidense
3	estadounidenses
3	ex
3	existente
3	existentes
3	exterior
3	extra
3	familiares
3	feito
3	feitos
3	finais
3	fiscal
3	fora
3	global
3	grandes
3	iguais
3	igual
3	impostos
3	independentes
3	individuais
3	industrial
3	insuficiente
3	interior
3	internacional
3	leve
3	locais
3	local
3	líder
3	líderes
3	maiores
3	mais
3	mal
3	material
3	medio
3	menor
3	menores
3	mil
3	minoristas
3	municipais
3	mínimo
3	mínimos
3	móbil
3	móviles

Nº	Forma
3	nacionalistas
3	naval
3	negativa
3	no
3	nós
3	occidental
3	off-shore
3	oitenta
3	optimista
3	optimistas
3	outra
3	outras
3	outro
3	outros
3	particular
3	pendentes
3	peor
3	perda
3	persoal
3	pesimista
3	pobres
3	populares
3	postos
3	precedentes
3	presentes
3	primeiro
3	principal
3	procedente
3	profissionais
3	profesional
3	propio
3	proprios
3	provinciais
3	quince
3	quinto
3	quén
3	reais
3	real
3	regular
3	residentes
3	responsable
3	seguro
3	sete
3	seus
3	social
3	socialista
3	suficientes
3	superior
3	superiores
3	só
3	súas
3	terceiro
3	titular
3	totais

Nº	Forma
3	trinta
3	variable
3	vostede
3	xa
3	xusto
4	1,2
4	10
4	11
4	12
4	120
4	138
4	15.000
4	19
4	2
4	21
4	22
4	25
4	3
4	300
4	300.000
4	32
4	35
4	4
4	400
4	5
4	500
4	51
4	55
4	60
4	63
4	65
4	66
4	75

Nº	Forma
4	85
4	9
4	Alicante
4	algo
4	bastante
4	ben
4	cada
4	cal
4	catorce
4	centrais
4	demais
4	editorial
4	final
4	industriais
4	lles
4	mellor
4	mesmo
4	moito
4	nada
4	nove
4	once
4	particulares
4	pouco
4	responsables
4	sesenta
4	seu
4	súa
4	tales
4	tanto
4	tarde
4	temporal
4	total
4	un

Nº	Forma
4	unha
4	vinte
4	web
5	15
5	20
5	200
5	30
5	40
5	50
5	80
5	catro
5	cen
5	cinco
5	dez
5	moral
5	nos
5	o
5	oito
5	presente
5	segundo
5	seis
5	sí
5	tal
6	100
6	calquera
6	me
6	tres
7	lle
8	quen
8	si
9	qué
12	que

Un aspecto fundamental a ter en conta nos datos deste subcorpus é a importante presenza de numerais, cuxo uso, esperable por ende, deriva precisamente da temática económica que centra os textos desambiguados. En concreto, das 21 981 unidades gramaticais distintas resultantes da etiquetación deses documentos, 2931, ou sexa o 13,33%, corresponden á categoría Numeral. Vexamos cal é o seu comportamento.

O sistema de numeración romano computa só sete análises gramaticais diferentes e emprégase residualmente neste subcorpus con dous usos ben diferenciados.

Por unha banda, como é habitual hoxe en día, utilízase para sinalar os séculos. Con esta función concorren cinco análises distintas (X, XVIII, XIX, XX, XXI) nas que caracterizamos o seu valor como cardinal, en concreto Ncnms (*Numeral, cardinal, non*

determinante, masculino, singular). Excepto nun caso no que está implícito, nos outros sempre precede o substantivo *século*. Algúns exemplos do seu uso son os seguintes:

Está documentado que no século X ata os bispos xantaban callos. [CORGAetq: CG2001-10-15/13]

O responsable de Loterías e Apostas do Estado (LAE) ultima estes días o Sorteo de Nadal, o último do século XX, pero coa mirada posta no XXI, no que se poderá comprar lotería tradicional por internet e cubrir tamén os boletos dos outros sorteos de Loterías e Apostas. [CORGAetq: CG2000-12-18/41]

Pola outra banda, os números romanos úsanse nos documentos deste bloque para facer referencia a algúns dos capítulos nos que se dividen os orzamentos autonómicos. Só dúas son as análises diferentes que contabilizamos neste apartado como Nonms (*Numeral, ordinal, non determinante, masculino, singular*) (VI, VII)¹³⁵. As distintas ocorrencias das dúas análises de números romanos con valor ordinal aparecen sempre referidas a capítulo. No seguinte exemplo reproducimos unha mostra das dúas análises mencionadas:

Se se fai a comparación sumando os capítulos VI e VII das Contas autonómicas _investimentos e transferencias de capital ó sector público e privado_ o resultado non é moito mellor, pois só dúas consellerías superan en recursos o gastado no Ano Santo. [CORGAetq: TN1998-04/12]

Como é natural hoxe en día na expresión dos números, o sistema de numeración arábigo é o que acolle o 99,76% das distintas análises de unidades pertencentes a este apartado, en concreto 2924. E tampouco debe sorprendernos, dados a temática e xénero do subcorpus, que 2757 desas análises se representen na escrita con cifra, fronte ás 167 nas que se recorre ás letras.

Das 43 análises distintas nas que os números árabes se caracterizan como ordinais, só unha ocorrencia aparece grafada con cifra:

Na súa intervención, Santiago Seco, director do Departamento de Análises, Prospeccións e Asesoramento, afondou no significado das ACIS e AXES, tanto para mellora-la calidade do leite coa vista posta no 1º de xaneiro do 98, no caso das primeiras, xa que a partir desa data non se poderá recoller leite que non reuna as condicións mínimas de asepsia segundo dispón a UE, coma para mellora-los resultados económicos das explotacións, coas segundas, e mesmo co servicio de laboratorio que se puxo en marcha. [CORGAetq: F1997-08/2]

¹³⁵ Para a súa caracterización véxase o apartado 1.3.2.7. Numeral (N).

No resto das 42 análises distintas cúmprese a observación realizada por González Rei (2004: 310) acerca de que os ordinais adoitan escribirse en letra ata o vixésimo, xa que neste subcorpus –non sendo o exemplo que vimos de ver– non hai ningún ordinal, nin en cifra nin en letra, que supere esa posición:

O xoves foi o décimo noveno día consecutivo en que a mostra de sete tipos de cru da organización _sobre a que se calcula o prezo medio do barril do cártel_ continuou cotizando moi por debaixo da banda de fluctuación de entre 22 e 28 dólares. [CORGAetq: CG2001-10-20/34]

A comunidade autónoma galega encabezou, entre os meses de xaneiro e agosto deste ano, as exportacións españolas de pescado, conservas, madeira, naval, pasta de madeira e rochas ornamentais, e foi a segunda comunidade exportadora de prendas e complementos de vestir, a terceira de automóbiles e leite, e a cuarta de aluminio. [CORGAetq: CG2000-11-26/13]

Caixavigo ocupa o sexto lugar entre as caixas de aforro e o vixésimo do ránking de entidades financeiras que operan en España, no apartado de mandatos e participacións de bonos. [CORGAetq: CG1996-02-29/41]

Polo que se refire ás 2881 análises distintas de unidades numéricas en arábigos caracterizadas como cardinais, 125 son representadas en letra e 2756 en cifra, o que supón unha porcentaxe de uso do 95,66% a favor das cifras. Este dato é esperable dado que estamos antes noticias xornalísticas cuxa temática específica se centra en calquera ámbito relacionado coa economía: bolsa, banca, facenda, empresa, orzamentos, subvencións, etc.

Non obstante, non só o tipo de texto inflúe na elección de cifras ou palabras na escritura dos números; a complexidade do número que se debe expresar e o contexto de uso tamén son determinantes na escolla. Así, tal e como describe González Rei (2004: 307-310), no subcorpus xornalístico de economía, o usual é escribir os cardinais en letras cando se trata das unidades e dos maiores de dez que constitúen palabra única:

Sete semanas máis tarde, Brasil estoupa de novo e o espectro do cesamento de pagos ronda. [CORGAetq: ANT2002-10-21/37]

En 1988, o país producía o vintecinco por cento dos buques que se construían en España. [CORGAetq: TN1997-04/7]

O mercado italiano continúa sendo o máis competitivo, o que lle permite ofrecer trinta modelos diferentes cos prezos (antes de impostos) máis reducidos da Unión Europea (UE). [CORGAetq: CG1996-02-16/108]

mais tamén ocorre o contrario; danse casos de unidades representadas en cifra e de números maiores de dez constituídos por unidades multipalabra escritos en letra:

Superávit da balanza por conta corrente no 95 de case 1 billón [CORGAetq: *CG1996-02-10/103*]

Neste momento, 4,2 millóns de familias contan con préstamos hipotecarios a interese variable, cunha débeda media viva que ascende a 8 millóns de pesetas. [CORGAetq: *CG2000-11-06/64*]

A sétima directiva comunitaria do ano oitenta e sete utilizou o argumento da crise como xustificación para manter a prohibición de que Astano producise buques. [CORGAetq: *TN1997-04/7*]

O saldo total de recursos propios de Caixavigo superou, ó peche do exercicio de 1995, os trinta e un mil millóns de pesetas, cifra que representa un incremento do 21% con respecto ó ano 94, indicou onte a entidade. [CORGAetq: *CG1996-02-03/48*]

Así mesmo, a busca de concisión e claridade, característica dos titulares, xustifica a aparición de dúas unidades numéricas en cifra en posición inicial absoluta:

140 millóns de marxe de explotación [CORGAetq: *F1997-04/7*]

400.000 contribuíntes utilizaron o programa "Padre" para pórse ó día con Facenda [CORGAetq: *CG1999-05-13/49*]

en contra do que dita a norma e do que é usual no resto do subcorpus, pois nas outras 25 ocasións en que atopamos un numeral en posición inicial absoluta este está grafado en letra:

Oito millóns de visitantes, dicíase dende a Xunta, "un gran éxito". [CORGAetq: *TN1998-04/12*]

Trinta mil millóns de dólares non lle dan a Brasil un colchón suficiente, declarou Roger Scher, analista da axencia inglesa Fitch, nunha das conferencias que seguiron ás xuntanzas do FMI e o Banco Mundial. [CORGAetq: *ANT2002-10-31/27*]

A complexidade do número que se expresa tamén condiciona a aparición da cifra, e así sempre que se presentan decimais ou a expresión numérica excede de tres palabras na súa representación con letras xorde a ortografía dos números con cifras. Algúns exemplos do exposto son os seguintes:

Este saldo estaba composto, o pasado venres, por 4,19 billóns de pesetas de bonos e obrigas do Estado adquiridos polos estranxeiros ata o vencemento dos títulos, así como por 40.800 millóns de pesetas de letras do Tesouro. [CORGAetq: *CG1996-02-17/95*]

Os gastos (939.218.967 pts.), cunha lixeira melloría (3.555.000), mantéñense nos mesmos niveis que en 1996. [CORGAetq: F1997-08/07]

3.2.2.2.3. *O subcorpus xornalístico xeral*

Sabedores de que o subcorpus de prensa económica desambiguado manualmente non era suficiente para servir de *gold standard* do xénero xornalístico, precisamente pola súa especificidade temática, decidimos ampliar o corpus de adestramento nunhas 150 000 palabras cun conxunto de xornais extraídos do CORGA. Este último bloque presenta a peculiaridade de que comprende todas as noticias de cada un dos xornais que nel se inclúen, representando polo tanto os diferentes estilos e características das distintas seccións, e non só da de Economía. Iniciamos esta segunda fase de desambiguación manual a mediados do 2008 e concluímos a finais do 2010, dando por finalizado o adestramento para o xénero xornalístico.

O subcorpus xeral de prensa está constituído por 596 arquivos en formato XML correspondentes a outras tantas noticias pertencentes ás seguintes publicacións:

Número 1166 do 3 de marzo do ano 2005 do semanario *A Nosa Terra*: 77 noticias

Número 1318 do 10 de xullo do ano 2008 do semanario *A Nosa Terra*: 89 noticias

Día 15 de maio de 2007 do xornal gratuito *De Luns a Venres*: 121 noticias

Día 17 de marzo de 2007 do xornal *Galicia Hoxe*: 89 noticias

Día 30 de marzo de 2007 do xornal *Galicia Hoxe*: 113 noticias

Día 5 de agosto de 2008 do xornal *Galicia Hoxe*: 107 noticias

Neste subcorpus analízanse 190 720 elementos gramaticais, dos cales constitúen análises diferentes 25 168 que implican a ocorrencia de 22 927 formas ou unidades gramaticais diferentes remisibles a 15 653 lemas distintos; así mesmo, nel aparecen documentadas 287 etiquetas das 383 posibles.

A distribución de formas por número de etiquetas correspondente ao módulo do subcorpus de prensa xenérico de XIADA é como segue:

Total de formas distintas:	22 927
----------------------------	--------

Total de formas+etiquetas+lemas diferentes: 25 167

Formas con 1 etiqueta:	21 160
Formas con 2 etiquetas:	1467
Formas con 3 etiquetas:	204
Formas con 4 etiquetas:	59
Formas con 5 etiquetas:	24
Formas con 6 etiquetas:	5
Formas con 7 etiquetas:	2
Formas con 8 etiquetas:	1
Formas con 9 etiquetas:	3
Formas con 10 etiquetas:	1
Formas con 16 etiquetas:	1

Se cos datos obtidos, seguindo a fórmula empregada por Graña (2000: 140), calculamos a porcentaxe de formas ambiguas e o número medio de etiquetas por forma, obtemos as seguintes cifras:

$$\% \text{ formas ambiguas} = \frac{\# \text{ formas ambiguas}}{\# \text{ formas}} \times 100 = \frac{22\ 927 - 21\ 160}{22\ 927} \times 100 = 7,70\%$$

$$\# \text{ nº medio de etiquetas por forma} = \frac{\# \text{ etiquetas}}{\# \text{ formas}} = \frac{25\ 167}{22\ 927} = 1,09 \text{ etiquetas por forma}$$

Tamén neste caso recollemos nos cadros seguintes as unidades léxicas do subcorpus xornalístico xeral que se correlacionan con tres ou máis etiquetas. A primeira das columnas acolle o número de etiquetas que se lle asignan neste subcorpus (Nº) e a segunda a forma ortográfica correspondente (Forma):

Nº	Forma
3	1,5
3	1,7
3	10
3	100
3	11
3	12
3	14
3	16
3	21
3	23
3	24
3	25
3	28
3	32
3	35
3	38
3	41
3	42
3	43
3	45
3	47
3	48
3	5
3	5.000
3	600
3	65
3	67
3	7
3	75
3	800
3	82
3	87
3	9
3	Breogán
3	EEUU
3	FEFN
3	Goya
3	Oscar
3	PXOM
3	Portos de Galicia
3	Rosalía de Castro
3	a
3	acompañante
3	acusado
3	afectada
3	afectados
3	alerta
3	algúns
3	antes
3	apartados
3	asistentes
3	ata
3	audiovisuais
3	axente

Nº	Forma
3	baixa
3	bastante
3	cando
3	canta
3	capaces
3	capaz
3	capital
3	cara
3	centrais
3	certo
3	chamada
3	cincuenta
3	civís
3	claro
3	como
3	convocantes
3	correspondente
3	corte
3	coñecidos
3	criminais
3	cuarto
3	cura
3	dependentes
3	descendente
3	despois
3	detidos
3	dezaoitto
3	diferentes
3	difficil
3	dito
3	dous
3	elegante
3	emigrante
3	era
3	escrita
3	escrito
3	especial
3	esta
3	estadounidense
3	estadounidenses
3	eu
3	ex
3	excelente
3	exterior
3	familiares
3	febles
3	feito
3	feitos
3	felices
3	feliz
3	fieis
3	finais
3	finalistas
3	fiscal

Nº	Forma
3	forestais
3	forestal
3	forte
3	fácil
3	gran
3	grave
3	habitual
3	igual
3	importante
3	imposible
3	independente
3	inmigrante
3	intelixente
3	interesante
3	interiores
3	internacionais
3	internacional
3	irregulares
3	locais
3	maior
3	maiores
3	mais
3	materiais
3	material
3	menos
3	min
3	mira
3	modelos
3	moitos
3	moral
3	mortais
3	multimedia
3	musical
3	móbil
3	naval
3	nin
3	non
3	nuclear
3	o
3	occidentais
3	oficiais
3	outra
3	outras
3	outros
3	paciente
3	paga
3	parcial
3	parte
3	particulares
3	pasado
3	pegada
3	pendente
3	persoal
3	personaxe

Nº	Forma
3	personaxes
3	pobres
3	política
3	popular
3	populares
3	posta
3	posto
3	previsíbel
3	primeiro
3	procedente
3	procedentes
3	produtora
3	profesional
3	propio
3	propios
3	representante
3	respecto
3	responsable
3	rock
3	rural
3	secuestrada
3	seguro
3	simple
3	surrealista
3	só
3	tales
3	tantos
3	te
3	terrorista
3	ti
3	tratados
3	universais
3	varias
3	verde
3	vista
3	vostede
3	vía
3	xa
3	xoves
3	xunta
3	xusto
3	árabe
4	13
4	15
4	150
4	17
4	18

Nº	Forma
4	2
4	20
4	3
4	40
4	6
4	70
4	8
4	80
4	Rosalía
4	algo
4	baixo
4	ben
4	cales
4	canto
4	dez mil
4	dobre
4	editorial
4	entón
4	final
4	galeguista
4	grandes
4	libre
4	lucense
4	mal
4	medio
4	mellor
4	menores
4	mesmo
4	militar
4	moito
4	nacionalista
4	occidental
4	once
4	optimista
4	peor
4	pouco
4	presente
4	presentes
4	profesionais
4	regular
4	responsables
4	responsábel
4	sete
4	seus
4	socialista
4	socialistas
4	son

Nº	Forma
4	súa
4	súas
4	titular
4	total
4	trinta
4	unha
4	web
5	26
5	4
5	50
5	60
5	cada
5	cal
5	catro
5	cen
5	central
5	cinco
5	doce
5	fora
5	máis
5	nacionalistas
5	nada
5	nove
5	nós
5	oito
5	protagonista
5	segundo
5	seu
5	tanto
5	tres
5	un
6	cantos
6	dez
6	lles
6	seis
6	tal
7	demais
7	si
8	nos
9	calquera
9	lle
9	me
10	quen
16	que

3.2.2.3. O tratamento das palabras descoñecidas

Con total seguridade, cando tratemos de etiquetar novos documentos, atopáremonos con palabras que non están introducidas no lexicón e que tampouco

presentan ningunha ocorrencia no corpus de adestramento. Isto implica que o etiquetador, *a priori*, non dispoñerá de ningunha información sobre a distribución das posibles etiquetas desas formas para as que, non obstante, debe aventurar unha análise. Pois ben, o procedemento de adiviñación e o tratamento das palabras descoñecidas que teña integrado o etiquetador pode marcar a diferenza no rendemento final.

Graña (2000: 95-98) dá conta de diversos traballos orientados a mellorar a estimación das probabilidades de emisión para as palabras descoñecidas dentro do marco dos modelos ocultos de Markov, e describe polo miúdo o método que actualmente ofrece mellores resultados para as linguas flexivas, que é o que está implementado no sistema XIADA: a análise de sufixos mediante aproximacións baseadas en inferencias Bayesianas. En termos xerais, segundo este método, para o cálculo da etiqueta dunha palabra descoñecida utilízanse as probabilidades de emisión das etiquetas para palabras que rematan co mesmo sufixo no corpus de adestramento, entendendo por sufixo a secuencia final de caracteres dunha palabra. Para o cálculo considérase só as palabras que aparecen menos de 10 veces e unha lonxitude máxima de sufixo de 10.

3.2.2.4. O corpus de aplicación vs. o corpus de adestramento e o lexicón

Tal e como viñemos indicando ao longo desta exposición, a nosa pretensión é aplicar paulatinamente o etiquetador automático estatístico do galego actual sobre o CORGA. En breve, toda a prensa do CORGA estará dispoñible en liña no *Corpus de Referencia do Galego Actual etiquetado automaticamente* (CORGAetqa) nunha versión etiquetada só automaticamente, sen posterior revisión manual. Así pois, as similitudes entre o corpus de adestramento e o corpus de aplicación, así como o feito de que o lexicón se incrementou, entre outras fontes, coas 100 000 formas máis frecuentes do CORGA, fannos pensar que a precisión vai ser alta.

A esta tese non contribúe só que os textos que se van etiquetar procedan da mesma fonte ou mostren un estilo semellante aos textos de adestramento, senón que está influenciada polo feito de que xa avaliamos o rendemento do etiquetador nun traballo previo.

A análise levada a cabo sobre a etiquetación das noticias do xornal *Galicia Hoxe* do día 17 de marzo de 2007 (Domínguez et al., 2008), onde verificamos un acerto do etiquetador do 95,99%, fannos prever, como mínimo, unha precisión do 96%, aínda que é posible que esta cifra se vexa incrementada por ter aumentado dende daquela o volume do corpus de adestramento, xa que a etiquetaxe do documento examinado só contou coa guía do *kernel* e do subcorpus de prensa económica.

3.3. A ambigüidade en galego: as anfiboloxías formais no corpus de adestramento

Xeralmente enténdese que unha palabra ou un enunciado son ambiguos cando presentan máis dunha interpretación; non obstante, nós non imos describir nesta epígrafe todos os tipos de ambigüidade existentes, senón só aqueles que implican un cambio na etiquetaxe, considerando nesta a cadea completa de análise de cada forma.

É certo que un problema importante no procesamento da linguaxe natural é determinar o sentido ou significado de palabras ambiguas nun determinado contexto e, tamén o é, que esta tarefa de desambiguación é complexa para a maioría das aplicacións do procesamento da linguaxe natural, mais o perfeccionamento na identificación do sentido correcto dunha palabra pode mellorar o comportamento de sistemas de tradución automática, sistemas de recuperación de información, etc., de aí que moitos autores dirixan os seus esforzos cara á desambiguación semántica automática (Ureña, 2001; Nica, 2006; Martínez, 2007, etc.).

O proceso de resolución da ambigüidade léxica consiste na identificación do significado correcto dun termo nun contexto dado de entre os proporcionados por un dicionario ou *thesaurus*, xa que unha mesma palabra pode ter diferentes acepcións dependendo do contexto do que se trate. A distinción de significados máis ou menos finos depende do dicionario que se empregue.

Dado que non incluimos información semántica nin na construción do etiquetario nin na do lexicón, e tampouco levamos a cabo o etiquetado semántico de corpus nin este forma parte dos nosos obxectivos, non imos tratar neste apartado a polisemia ou ambigüidade producida polas distintas acepcións dun mesmo lema, posto que a ampliación de significados non implica alteracións nas súas características morfolóxicas e, polo tanto, a estes efectos non repercute na actuación do etiquetador.

Pensemos no substantivo *banco* e tomémolo como mostra de forma polisémica. Segundo o DRAG (1997) posúe as seguintes acepcións:

banco *s.m.* **1.** Asento de madeira, pedra ou outro material, con ou sen respaldo, onde se poden sentar varias persoas a un tempo. *Sentámonos nun banco do xardín.* **2.** Entidade financeira, estatal ou privada, que se dedica a comerciar co diñeiro que recibe en depósito dos seus clientes, á concesión de créditos, etc. *Ingresei os cartos no banco.* **3.** Mesa de traballo dalgúns artesáns. *Un banco de carpinteiro.* **4.** Acumulación de area ou grava no leito dun río ou no litoral, que ocasiona a elevación do fondo e constitúe un perigo para a navegación. *O buque encallou nun banco de area.* CF. **baixío.**

Porén, no noso lexicón non se recollen as distintas acepcións e só figura unha entrada *banco*, caracterizada como *Substantivo masculino*. Isto implica que *banco*, independentemente do sentido que denote, vaise etiquetar sempre como substantivo masculino, polo que, para o etiquetador e polo que a nós respecta neste estudo, non é unha forma ambigua. Así, facendo unha revisión manual dos 79 casos desta forma que aparecen na versión 2.4 do CORGAetq vemos que só dous responden á primeira acepción e os setenta e sete restantes á segunda; mais para obter este dato tivemos que valernos dunha comprobación manual de todos os resultados da busca, xa que o tratamento morfosintáctico que reciben estas ocorrencias no CORGAetq é idéntico en todos os casos: Scms (*Substantivo común masculino singular*); nada hai no sistema de XIADA ou no corpus que nolas individualice por acepcións.

O preparador deportivista deu as primeiras pistas da aliñación que presentará fronte ós cataláns, ante os que, consonte co visto na última sesión preparatoria, deixaría no banco a Riki _a fichaxe máis cara do Dépor esta temporada_ e metería no seu lugar o ex xogador do conxunto catalán, Joan Verdú, na media punta, mentres que o asturiano Adrián López, que foi titular no último encontro de Liga ante o Getafe, sería a referencia ofensiva dos coruñeses. [CORGAetq: GH2007-03-30/73]

O banco emisor europeo xustificou a compra de euros contra dólares e iéns coa necesidade de previr unha ulterior caída da divisa europea e as presións inflacionistas na eurozona. [CORGAetq: CG2000-11-07/78]

Tampouco ten sentido incluír nesta descrición da ambigüidade os casos de homonimia homófona pero non homógrafa, pois ao non haber coincidencia na escrita, por moito que dúas formas se pronuncien do mesmo xeito, non vai ter repercusións na etiquetaxe de corpus, xa que se grafan de maneiras distintas e corresponden etiquetas diferentes. É o caso, por exemplo, de *ai* e *hai*, ambos presentes no CORGAetq e para o

que as análises unívocas son respectivamente, Y (*Interxección*) e Vpi30s (*Verbo, presente de indicativo, terceira persoa, xénero non aplica, singular*):

As comunidades máis mediterráneas e Madrid son, ¡ai!, contribuíntes netos. [CORGAetq: CG2002-05-04/16]

Pero hai realidades na pesca que foron a peor por mor da integración en Europa, di o conselleiro delegado de Pescanova. [CORGAetq: TN1997-04/7]

Cuestión moi distinta son para nós os casos de homonimia homógrafa, aínda que non se dea tamén a homófona, posto que sempre provocan a aparición de máis dunha etiqueta e, a miúdo, incluso ademais a posibilidade de máis dunha segmentación.

Pensemos, por exemplo, nos verbos da segunda conxugación que posúen como vogal radical *e* ou *o*. En todos eles se produce a coincidencia formal, que non fonética, entre a terceira persoa do presente de indicativo e a segunda do singular do imperativo. Exs.: *el t[E]n / t[e]n ti; el c[O]se / c[o]se ti; el c[O]me / c[o]me ti; el b[E]be / b[e]be ti*, etc.

Semellantes aos casos anteriores son, por exemplo, *destes, polos* e *velas*, co agravante de que nestes, a maiores, se orixina unha segmentación diferente da unidade dependendo da análise que corresponda. Así, coa vogal tónica pechada corresponden

destes:

de [Preposición] + estes [demostrativo]

polos:

por [preposición] + los [artigo]

polos [substantivo]

velas:

ver [infinitivo *ver*] + las [pronome de acusativo]

ves [2ª singular do presente de indicativo de *ver*] + las [pronome de acusativo]

mentres que coa aberta xa estamos ante:

destes:

destes [2ª plural do pretérito de indicativo de *dar*]

polos:

pos [2ª singular do presente de indicativo de *pór*] + los [pronome de acusativo]

pos [2ª singular do presente de indicativo de *poñer*] + los [pronome de acusativo]

velas

velas [substantivo]

velas [2ª singular do presente de indicativo de *velar*]

Como falantes non podemos consideralas formas homónimas, pois distinguimos perfectamente entre elas polo grado de abertura da vogal. E se no corpus procederamos a realizar unha trascrición fonética, tampouco consideraríamos nelas a existencia de ambigüidade; mais, xa que nos limitamos á escrita e que non se marca de ningún xeito a abertura das vogais, atopámonos con que un mesmo *token* recibe máis dunha análise, o que incrementa a ambigüidade e as posibilidades de erro na etiquetación automática.

A continuación veremos, pois, que tipos de ambigüidade se dan no nivel morfosintáctico do noso corpus e de que maneira estes repercuten na etiquetaxe.

3.3.1. Tipos de ambigüidade

En liñas xerais, distinguimos os seguintes catro tipos de ambigüidades no nivel morfosintáctico.

3.3.1.1. Homonimia intercategorial ou ambigüidade entre categorías.

Caracterízanse neste tipo as formas que son atribuíbles a categorías gramaticais distintas. É o que ocorre, por exemplo, cos *tokens baixo, a* ou *que*:

<u>baixo</u> :	<u>a</u> :	<u>que</u> :
Adxectivo	Preposición	Conxunción
Preposición	Artigo	Relativo
Substantivo	Pronome	Interrogativo-exclamativo
Adverbio	Substantivo	Substantivo
Verbo		

Este tipo de anfibloxía desfaiase sempre na práctica da etiquetaxe de corpus, sexa esta manual ou automática, pois o contexto sempre nos dá a pauta para poder diferenciar entre as distintas categorías e escoller a que corresponda. Vexamos un exemplo real de uso de cada unha das categorías relacionadas para o primeiro dos termos, recollidos segundo a orde categorial anterior:

Non obstante, o director xeral de Recursos Humanos do Sergas, Antonio Fernández Paniagua, considerou "baixo" o seguimento da xornada de paro. [CORGAetq: GH2007-03-30/8]

Baixo esta hipótese, Royal Bank of Scotland saíra das quinielas do Santander Central Hispano. [CORGAetq: CG2002-05-04/5]

Acordeón, guitarras, baixo e percusión complementan un proxecto que non pretende unha formación de moitos músicos e que procura unha sonoridade diferente, máis limpa e cun certo carácter minimalista. [CORGAetq: ANT2005-03-03/69]

Escribir e ler seguen algo por baixo, porque houbo xeracións sen galego na escola. [CORGAetq: ANT2008-07-10/15]

Aurelio: Cando ti sobes eu baixo e ó revés. [CORGA: A.Vila e X. Cid cabido; *Copenhague. Comedia evidencialista*, Xerais, 1993]

3.3.1.2. Homonimia intracategorial ou ambigüidade nos atributos.

Os *tokens*, a pesar de clasificarse no interior dunha mesma categoría gramatical, difiren nalgún ou en varios dos atributos. Dentro deste apartado podemos establecer varios subgrupos en función da súa resolución:

1) Ambigüidade atributiva só teórica

A ambigüidade é só teórica, pois na práctica da etiquetaxe de corpus sempre corresponde un dos valores e é imposible mantela. Dáse esta situación por exemplo, entre outros, por unha banda cos atributos *determinante*, *non determinante* nos posesivos, demostrativos, numerais, relativos e interrogativos-exclamativos, e por outra co valor *nominativo* e *preposicional* nos pronomes tónicos, posto que en ambos os casos a presenza dun dos valores invalida a existencia do outro, e o etiquetador, obrigatoriamente, ten que optar por unha das dúas posibilidades.

Exemplos:

Pero estes non son os únicos que se pagan. [CORGAetq: CG2000-11-11/43]

Barón de Ley é das poucas compañías que nestes intres acumulan beneficios no parque ao longo do ano. [CORGAetq: ANT2002-10-17/29]

Hai que destacar que non se esixe fianza de ningún tipo para asistir á poxa nin para poder poxar e participar nela. [CORGAetq: CG1999-05-14/62]

En realidade, os médicos víronlle dous DIU, un ben posto, que ela mesma se colocou, "e outro no corpo". [CORGAetq: GH2007-03-17/41]

2) Ambigüidade atributiva ocasional

A ambigüidade débese a unha cuestión de inespecificación e non sempre é posible determinar que valor procede. O caso máis evidente é o dos adxectivos invariables no xénero, tipo *agradable*, que no lexicón levan unha caracterización tripla neste atributo: *masculino*, *feminino* e finalmente *masculino/feminino*, para o caso de que os elementos presentes no contexto non nos permitan a correspondencia cun dos dous xéneros.

Na seguinte secuencia, por exemplo, a presenza dos adxectivos masculinos *fermoso*, *decorado* e *limpo* indícanos que *agradable* referencia tamén o xénero masculino (A0ms):

Este quedou fermoso, ben decorado, agradable e limpo. [CORGAetq: GH2007-03-30/112]

Pola contra, nestoutra o funcionamento como modificador dun substantivo feminino lévanos á caracterización feminina de *amable*:

O segundo ano do bienio fiscalizado presenta unha cara máis amable, xa que conclúe cun beneficio de 2.400 millóns de pesetas. [CORGAetq: TN1998-04/12]

Por último, a lectura dalgúns dos símbolos mediante os que se acostuma valorar unha obra teatral, un libro ou un complexo residencial provoca que non teñamos un referente claro e de aí a caracterización co valor ambiguo *masculino/feminino* nos catro casos localizados no GH2007-03-30/102 do CORGAetq:

*Regular

**Interesante

***Notable

****Excelente

Mais non é preciso ser tan rebuscados para atopar outro exemplo deste tipo de ambigüidade; a brevidade e concisión típica dos titulares provoca que en ocasións non poidamos atribuír claramente un dos dous xéneros e nos vexamos obrigados a manter a caracterización co valor ambiguo, deixando este sen especificar:

"Inaccesible" para o 70 por cento dos cidadáns [CORGAetq: CG2000-12-29/88]

Un exemplo máis deste subtipo de anfiboloxías ten lugar nos paradigmas verbais das tres conxugacións na primeira persoa do plural entre o presente e o pretérito de

indicativo, excepto, por suposto, nos dezaioito verbos que presentan un tema de perfecto distinto do tema de presente¹³⁶.

Na maioría das ocasións, o contexto permítenos dilucidar se a forma verbal refire un tempo presente ou un pasado, debido á presenza doutras formas verbais ou de adverbios que nos axudan a situar a referencia temporal.

Así, na seguinte secuencia, os pretéritos *fixo* e *deu*, xunto coa mención do ano 1986, permítennos caracterizar o token *chegamos* como pretérito:

Ese traballo fixose no referendo do 1986 e deu froito alí onde chegamos a falar coa xente, de xeito directo. [CORGAetq: ANT2005-03-03/63]

Pola contra, a presenza de formas verbais de presente, tanto de indicativo (*é*) como de subxuntivo (*continúe*), indicánnos que a forma *agardamos* refire o presente:

O que si agardamos é que continúe habendo amor para que o conto continúe, aínda que xa se sabe que cos anos e coa axenda real os detalles déixanse un pouco de lado. [CORGAetq: DLV2007-05-15/39]

Non obstante, noutras ocasións o contexto oracional non nos proporciona información suficiente para tomar unha decisión na asignación dun valor ou outro. Nestes casos podemos falar de ambigüidades xenuinamente textuais, mais, dado que no deseño do etiquetario non tivemos en conta a posibilidade de non diferenciar entre estes dous tempos verbais, deben ser resoltas, á forza, a favor dun ou doutro valor. Quizais inconscientemente cando se dá esta circunstancia deixámonos influír polo texto en xeral ou por apreciacións extralingüísticas e non nos limitamos só ao contexto oracional que rodea a forma en cuestión. Unicamente deste xeito podemos xustificar a escolla de pretérito para a primeira das seguintes secuencias e a elección de presente para a segunda delas:

Aumentamos o catálogo de novidades ao tempo que cambiamos o deseño para darlle un ton máis actual. [CORGAetq: ANT2008-07-10/5]

O titular de Interior engadiu que con estes proxectos "ampliamos os dereitos dos gardas civís e ampliamos a protección dos dereitos dos cidadáns". [CORGAetq: GH2007-03-17/66]

¹³⁶ Trátase dos verbos *dar*, *estar*, *ver*, *dicir*, *facer*, *querer*, *traer*, *caber*, *haber*, *saber*, *pracer*, *poder*, *ter*, *ser*, *ir*, *poñer*, *pór* e *vir*.

3.3.1.3. Homonimia categorial e atributiva plena

Son poucos os casos que se contabilizan neste tipo de ambigüidade, xa que a súa principal característica consiste en que unha forma presenta dúas análises nas que recibe a mesma etiqueta, coincidindo neste sentido non só a categoría na que se clasifica o *token*, senón tamén todos os atributos, só que cada unha das análises remítese a lemas diferentes. Prodúcese esta situación nas formas verbais do tema de perfecto dos verbos *ser* e *ir*, nas cales só o lema nos axuda a diferencialas.

Naturalmente, o contexto permítenos desfacer sempre a ambigüidade e escoller a análise correcta, mais, paradoxalmente, o etiquetador, por non ter nunca en conta a información relativa aos lemas, escolle un deles ao chou. Nunhas ocasións atina e noutras equivócase; véxanse senón as escollas que realiza o etiquetador de XIADA¹³⁷ para a forma *fun* nas secuencias seguintes, clarísimas para unha persoa pero problemáticas para unha ferramenta que ignora os lemas:

Oración:	Onte fun a Vigo.				
Formas:	onte	fun	a	Vigo	.
Etiquetas:	Wn	Vei10s	P	Sp00	Q.
Lemas:	onte	ser	a	Vigo	.

Oración:	Onte fun a mellor das dúas.							
Formas:	onte	fun	a	mellor	de	as	dúas	.
Etiquetas:	Wn	Vei10s	Ddfs	Acfs	P	Ddfp	Ncnfp	Q.
Lemas:	onte	ir	o	mellor	de	o	dous	.

Oración:	Onte fun a unha única clase.						
Formas:	onte	fun	a	unha	única	clase	.
Etiquetas:	Wn	Vei10s	P	Difs	A0fs	Scfs	Q.
Lemas:	onte	ir	a	un	único	clase	.

¹³⁷<http://corpus.cirp.es/xiada/etiquetador.html> na súa versión 2.4 [Executado o 27/01/2011].

Un caso moi peculiar deste tipo de ambigüidade é o que se dá nas formas verbais comúns aos verbos *poñer* e *pór* e aos compostos de ambos (todo o tema de pasado, o presente de subxuntivo, o participio, as tres persoas do singular do presente de indicativo e a segunda de singular do imperativo), sinónimos en todas as súas acepcións (VOLGa, 2004; Ir Indo, 2004; Xerais, 2009). Nestas circunstancias a ambigüidade é imposible de desfacer incluso para un lingüista, posto que, a non ser que na mesma oración se empreguen algunhas das formas verbais nas que os dous paradigmas difiren (copretérito, futuro e pospretérito de indicativo; infinitivo, infinitivo conxugado, xerundio, as tres persoas plurais do presente de indicativo e a segunda plural do imperativo) –co que saberíamos a que lema remitir a entrada–, dende o punto de vista normativo os dous lemas son correctos e, aínda que xeograficamente a extensión de *pór* é algo menor ca a de *poñer*, non imos facer pescudas extralingüísticas para saber onde naceu un determinado autor e comprobar se pertence a unha zona de *pór* ou de *poñer*. A solución menos comprometida ao noso ver nas formas verbais comúns a ambos, unha vez descartada a posibilidade dun lema dobre, seguindo o que se declara nas *Normas* (RAG/ILG, 1995: 157; 2003: 158), é darlle preferencia a *poñer*, considerándoo o termo non marcado do par verbal.

Finalmente, á parte das anfiboloxías xa comentadas dos verbos *ser/ir* e *pór/poñer*, no corpus de adestramento preséntansenos outros casos de coincidencia total entre forma e etiqueta cuxo lema é o que establece a diferenza. Dado que o seu número non chega á decena, imos recollelos todos, intentando agrupalos.

As dúas formas que encabezan esta nosa particular lista son *heroína* e *raíña*, ambas caracterizadas como Scfs (*Substantivo, común, feminino, singular*). Nas dúas, o lema é o que xustifica a existencia de dúas formas diferenciadas.

Así, nun caso *heroína* forma parte do paradigma de *heroe* e no outro é *heroína*, o estupefaciente, o representante do seu grupo:

heroína Scfs, lema *heroe*

A aparición de ETA fixo despois que Irlanda seguise de referencia, con simpatías para Bernardete Devlin, un pouco heroína como David fronte a Goliat. [CORGAetq: GH2007-03-30/88]

heroína Scfs, lema *heroína*

Incautado con seis doses de heroína en Camariñas [CORGAetq: GH2008-08-05/19]

Algo semellante ocorre con *raíña*, que aparece como feminino de *rei* e como sinónimo de *raiola*, con lema propio¹³⁸:

raíña Scfs, lema *rei*

Un audaz financeiro, José de Salamanca y Mayol, realizou o proxecto de creación dun organismo que levaría o nome da raíña Isabel II. [CORGAetq: CG1996-02-09/57]

raíña Scfs, lema *raíña*

Marzo, marzán, cara de can: pola mañá sol da raíña e pola noite afeitán as barbas cunha fouciña. [CORGAetq: GH2007-03-17/79]

No seguinte par o uso do castelanismo *xoves*, en vez de *mozos*, provoca a coincidencia formal e categórica plena con *xoves*, o día da semana:

xoves Scmp, lema *xoves*

A partir desta semana, Cuatro dedicará a noite dos xoves ao xénero policíaco. [CORGAetq: GH2008-08-05/103]

xoves Scmp, lema *xove*

Despois de varios problemas políticos que lle causaron dúas fuxidas de España, unha cara a Suíza e logo, xa en 1976, despois da morte de Franco pero antes do goberno democrático, cara á súa terra natal mexicana, chega agora cun grupo formado por estudantes e traballadores dos municipios de San Pedro das Colonias, Francisco I. Madero e Torreón Coahuilla que segundo di "pretende potenciar a inqueda, entre xoves e adultos como futuros espectadores ou artistas potenciais". [CORGAetq: GH2007-03-30/39]

A seguinte parella está protagonizada polo forma *total*, caracterizada como adverbio nuclear en ambos os casos e diferenciados á súa vez cada un deles polo lema:

total Wn, lema *total*

o diálogo, do que entendemos só a parte inglesa (por certo, dobrada, cando puideron deixala xa tal e como estaba, total para catro palabras e media...) explícanos máis da Revolución Cultural e da Revolución dos Aiatolas que trescentos artigos de opinión periodísticos. [CORGAetq: ANT2008-07-10/44]

total Wn, lema *totalmente*

¹³⁸ Se concorrese como forma diminutiva de *raia*, atopariámonos cunha ambigüidade tripla, pois o lema neste último caso sería a forma base *raia*: *raíña* Scfs, lema *raia*.

O ministro de Economía e Facenda, Pedro Solbes, afirmou en declaracións que sería un erro que Argentaria desaparecese e pasase a formar parte, total ou parcialmente, doutras entidades financeiras. [CORGAetq: CG1996-02-12/98]

Na primeira secuencia a clasificación nos dicionarios de *total* como lema é incuestionable, así como tamén o é na segunda que ante unha coordinación entre adverbios en *-mente* os manuais de estilo marcan que os anteriores á conxunción aparezan sen o sufixo. Nós cremos que nunha coordinación de adverbios en *-mente*, como a do último exemplo, todos os membros se deben remitir ao lema en *-mente*, posto que esa variante gráfica é só un alomorfo en distribución complementaria.

A ambigüidade nas restantes formas que nos quedan por ver deste subapartado é en certa medida responsabilidade nosa, xa que se orixina polo tratamento que lle outorgamos a un dos membros dos pares. Por que afirmamos isto? Porque o trato que reciben pares como *produtor/produtora*, *político/política*, por unha banda, e, pola outra, *costa/costas* e *táboa/táboas* diverxe nos dicionarios galegos que consultamos.

Así, no Xerais, tanto na súa edición do 2000 como na do 2009, dos pares anteriores só discriminan en entradas diferentes *política* e *político-a*; o DRAG (1997, versión provisional da nova edición¹³⁹ e mais a nova edición xa en liña¹⁴⁰) non discrimina en entradas diferentes ningún membro dos pares, pois remite *política* e *produtora* respectivamente a *político -a* e *produtor -ora*; finalmente, no Ir Indo, excepto para *costa/costas* e *táboa/táboas*, distínguese baixo entradas diferenciadas *política* e *produtora*.

Nós consideramos, cremos que acertadamente dende o punto de vista lingüístico, que deberían conformar entradas distintas no lexicón, e en consecuencia actuar como lemas diferentes na etiquetaxe de corpus, xa que en todos os pares se produce unha variación con respecto ás características dos membros do paradigma do lema canónico; isto é, ten lugar unha redución no paradigma en relación coa clasificación como substantivo con variación de xénero e número e como substantivo feminino con variación de número, posto que só parte da entrada se adecúa ás características gramaticais do lema. Por exemplo, baixo o lema *político* da clase substantivo só se sitúan as persoas que se dedican aos asuntos de goberno ou se interesan por eles; baixo

¹³⁹ Consulta realizada o 1 de febreiro de 2011 ao Seminario de Lexicografía da RAG.

¹⁴⁰ Consulta realizada o 10 de outubro de 2012 en <http://www.realacademiagallega.org/diccionario>.

este lema residen catro formas: *político*, *políticos*, *política* e *políticas*. Agora ben, baixo a clasificación de substantivo feminino, entre outras coa acepción de doutrina referente ao goberno dunha comunidade, non teñen cabida os masculinos *político*, *políticos*; se nesta acepción sempre é substantivo feminino non podemos polo tanto catalogar as súas formas dentro do paradigma de *político*.

Algo semellante ocorre con *costas* e *táboas*. Calquera das dúas en varias das súas acepcións¹⁴¹ son substantivos femininos nos que se produce variación de número; non obstante, nas acepcións de *parte posterior do corpo humano* ou *estilo de natación e empate, situación pola que ningún dos xogadores pode gañar a partida* non existe oposición numérica, pois con eses valores só poden concorrer na súa forma plural. De aí que desagreguemos en dous lemas cada un dos pares e lles outorguemos ese estatus a formas que na tradición lexicográfica non o teñen: *editora*, *construtora*, *política*, *produtora*, *costas*, *táboas*, etc.

Vexamos agora algúns dos exemplos que deron pé á problemática anterior.

política Scfs, lema *político*

A política viguesa pensa que o encontro supuxo un triunfo para Galicia pero tamén para as mulleres que "levan moito tempo gañando no PSOE". [CORGAetq: ANT2008-07-10/75]

política Scfs, lema *política*

Isto supuxo un avance moi grande na política económica española e desembocou nunha frenética actividade financeira. [CORGAetq: CG1996-02-09/57]

produtora Scfs, lema *produtor*

Chelo Loureiro, produtora executiva de Continental Producciones, reflexiona sobre o encontro. [CORGAetq: ANT2008-07-10/23]

produtora Scfs, lema *produtora*

Telefónica desfaise da produtora Endemol [CORGAetq: DLV2007-05-15/96]

costas Scfp, lema *costa*

¹⁴¹ *Costa*: s.f. 1. Zona de terra que está próxima ó mar, así como a zona do mar próxima á terra. 2. Terreo en pendente, particularmente referido a un camiño ou outra vía. *Táboa*: s.f. 1. Peza de madeira plana e pouco grosa, polo xeral de forma rectangular e bastante máis longa ca larga. 2. Por ext., peza plana e pouco grosa de certos materiais rixidos. 3. Cadro de números ordenados conforme a un método para facilita-la obtención duns determinados valores. 4. Relación ordenada e sistemática de determinadas materias, que permite consultalas ou localizalas rapidamente. [DRAG, 1997]

O sindicato aproveitou así a operación dada a coñecer onte, de detención dun veleiro con 3.800 quilos de haxix fronte ás costas da Coruña para pór de manifesto as condicións nas que realizan o seu traballo os efectivos deste corpo. [CORGAetq: GH2008-08-05/11]

costas Scfp, lema *costas*

Nun momento no que a industria do disco lle vira as costas á música que non é comercial, sacar un disco para o grupo é un proxecto "que nos fai ilusión pero que aínda está lonxe".[CORGAetq: ANT2005-03-03/69]

táboas Scfp, lema *táboa*

Na oficina da Autoridade Rexional os danos humanos e materiais convértense en números nas táboas e en imaxes nas fotos que falan dos labores de rescate, da recuperación e incineración de corpos, do restablecemento das comunicacións. [CORGAetq: ANT2005-03-03/67]

táboas Scfp, lema *táboas*

A loita quedou en táboas coa adopción da fórmula "lingua nacional e unificadora" para o inglés. [CORGAetq: GH2007-03-17/84]

Se partimos do número de análises distintas que se dan no corpus de adestramento de XIADA, 38 497, a incidencia deste apartado supón un 0,2% na problemática da multiplicidade de análises para a escolla do etiquetador automático. Agora ben, se o que temos en conta é a totalidade do corpus de adestramento, os 431 574 elementos gramaticais que se analizan, sumando todas as ocorrencias das sete unidades deste subapartado¹⁴², chegamos ás 370, o que supón un insignificante 0,08%.

3.3.1.4. Homonimia segmental ou ambigüidade na identificación do *token*.

Unha das primeiras tarefas que realiza o etiquetador/lematizador, cando se executa, consiste na delimitación das distintas unidades léxicas do texto que posteriormente será etiquetado. É unha das tarefas máis complexas e, ao tempo, decisiva, pois unha parte importante da taxa de acerto do programa vén determinada pola correcta identificación e segmentación dos elementos a etiquetar.

¹⁴² Non incluímos no cómputo final a forma *total* porque a análise como adverbio cuxo lema é *totalmente* introducímola á man, e, polo tanto, para o etiquetador non é unha posible opción. De momento, o etiquetador non ten en conta nunca como adverbio aquel que estando coordinado a outro en *-mente* non presenta tamén ese sufixo, porque esa información non a ten nos dicionarios cos que traballa. A nosa idea é implementar proximamente no sistema unha regra lingüística segundo a que, se unha forma é posible analizala como A0fs (*Adxectivo, grao non aplica, feminino, singular*), está seguida da conxunción *e* ou *ou* e segue un adverbio rematado en *-mente*, o etiquetador impida a caracterización inicial de adxectivo feminino e outorgue a de adverbio, introducindo no lema a forma que analiza rematada en *-mente*.

O problema prodúcese porque á hora de anotar un texto non podemos identificar unidade léxica con forma ortográfica. Por unha banda, está a existencia de secuencias de palabras que nun contexto dado se consideran unidades multipalabra, como son moitos nomes propios ou os distintos tipos de locucións recoñecidos, pero que nun contexto diferente deben tratarse como unidades diferenciadas. Por outra banda, está a presenza de formas gráficas que, en función do contexto, poden interpretarse como unha única unidade léxica ou ser segmentable en varias.

Así pois, dous son os tipos de homonimia segmental que consideramos para o galego:

- 1) aquela que se produce a nivel de palabra e que ten a súa orixe na existencia de amálgamas, por exemplo *destes* [contracción de preposición e mais demostrativo, fronte a forma verbal], *polos* [substantivo, fronte á contracción de preposición e mais artigo, e fronte á amálgama de forma verbal con pronome enclítico], *á* [substantivo, fronte a contracción], *selos* [substantivo, fronte a forma verbal con enclítico], etc.
- 2) aquela que ten lugar cando unha secuencia de formas gráficas é susceptible de ser interpretada como unha única unidade multipalabra ou ben como tantas unidades gráficas posúa a secuencia. Dáse este tipo, sobre todo, cos propios e coas locucións, mais tamén coa combinación doutros tipos de palabra.

Pensemos, por exemplo, na secuencia *Foron vinte e tres máis tarde*. Deixando de lado que a ambigüidade é irresoluble se atendemos só á secuencia mesma¹⁴³, no texto etiquetado automaticamente aparece nada máis unha das dúas opcións seguintes, correspondentes a cadansúa segmentación:

- 1) <unidade>vinte e tres</unidade>
- 2) <unidade>vinte</unidade>
- <unidade>e</unidade>
- <unidade>tres</unidade>

¹⁴³ É unha unidade (*vinte e tres*) ou son tres (*vinte / e / tres*)?

Este tipo é moi difícil de avaliar tanto en termos cualitativos coma cuantitativos, posto que o etiquetador xa opta por unha das posibles segmentacións e só ofrece a análise para a segmentación elixida. Amais, para poder levar a cabo dito estudo, a desambiguación tería que realizarse con esa única finalidade, a de detectar e anotar os casos de ambigüidade segmental manualmente, o que non compensaría, dado que o proveito que se lle podería sacar a tal estudo sería mínimo en relación co esforzo invertido en tal tarefa¹⁴⁴.

Cuestión distinta é o tipo de ambigüidade segmental primeiro, o 1). Malia que non podemos ofrecer cifras concretas do corpus de adestramento, pois unha vez que os textos son desambiguados elimínanse automaticamente todas as alternativas non válidas e non conservamos copia das fases previas, podemos, iso si, facer unha pequiñiña cala: coller as 1000 formas máis frecuentes do CORGA e etiquetalas, fixándonos en cales presentan máis dunha alternativa. A consulta ao arquivo etiquetado especificamente con este propósito amósanos alternativas segmentais diferentes para vinte e cinco unidades ortográficas. Recollémolas deseguido nunha táboa, na que a primeira columna inclúe a forma ortográfica e, as seguintes, as distintas alternativas de análise posibles para cada unha das unidades. Incluímos primeiramente a alternativa correspondente á unidade sen segmentar e a continuación as alternativas relacionadas con distintos tipos de amálgamas:

¹⁴⁴ Pénsese que as tarefas de desambiguación son xa bastante complexas como para requirir toda a atención, co que sería moi probable que se producisen erros, ben na recollida das unidades con ambigüidade segmental ben na mesma revisión da etiquetaxe.

Forma ortográfica	Alternativa 1	Alternativa 2	Alternativa 3
á	Substantivo	Contracción de preposición e artigo	
ó	Interxección	Contracción de preposición e artigo	
nos	Pronome	Contracción de preposición e artigo	
polo	Substantivo	Contracción de preposición e artigo	Amálgama de forma verbal e clítico
pola	Substantivo	Contracción de preposición e artigo	Amálgama de forma verbal e clítico
coa	Verbo	Contracción de preposición e artigo	
ás	Substantivo	Contracción de preposición e artigo	
día	Substantivo	Amálgama de forma verbal e clítico	
ela	Pronome	Amálgama de forma verbal e clítico	
días	Substantivo	Amálgama de forma verbal e clítico	
polos	Substantivo	Contracción de preposición e artigo	Amálgama de forma verbal e clítico
coas	Verbo	Contracción de preposición e artigo	
polas	Substantivo	Contracción de preposición e artigo	Amálgama de forma verbal e clítico
daquela	Adverbio Conxunción	Contracción de preposición e demostrativo	
ca	Conxunción	Contracción de preposición e artigo	
vello	Adxectivo Substantivo	Amálgama de forma verbal e clítico	
estamos	Verbo	Amálgama de forma verbal e clítico de acusativo e mais dativo	
río	Substantivo Verbo	Amálgama de forma verbal e clítico	
creo	Verbo	Amálgama de forma verbal e clítico	
dese	Verbo	Contracción de preposición e demostrativo	
fala	Substantivo Verbo	Amálgama de forma verbal e clítico	
seguinte	Adxectivo Substantivo	Amálgama de forma verbal e clítico	
vella	Adxectivo Substantivo	Amálgama de forma verbal e clítico	
elas	Pronome	Amálgama de forma verbal e clítico	
destes	Verbo	Contracción de preposición e demostrativo	

Na análise das 1000 formas máis frecuentes do CORGA xorden aínda outras unidades con ambigüidade segmental que partillamos das anteriores, porque en certa medida énos atribuíble a súa existencia. No interior destas últimas diferenciamos tamén dous grupos en función do motivo que provoca a ambigüidade.

O primeiro dos grupos orixínase por non estar restrinxida, en xeral, a combinación de clíticos con formas verbais¹⁴⁵. Trátase de tres formas do futuro de subxuntivo; mais é unha ambigüidade falsa por canto é moi improbable que estas formas poidan levar enclíticos: non aparecen en cláusulas independentes, polo que, en caso de habelos, o normal é a anteposición e non a énclise.

Unidade ortográfica	Alternativa 1	Alternativa 2
forma	Substantivo Verbo	Amálgama de forma verbal e clítico de dativo e mais de acusativo
vila	Substantivo	Amálgama de forma verbal e clítico
forte	Adverbio Adxectivo Substantivo	Amálgama de forma verbal e clítico

O segundo dos grupos xorde por dar cabida á etiquetaxe de formas non normativas ou incorrectas. É o caso de *da*, que aparece en bastantes ocasións como forma verbal de *dar*, o que leva á ambigüidade coa contracción de preposición e artigo; o do adverbio *no*, que coincide formalmente coa contracción de preposición e artigo; e, finalmente, tamén as contraccións *dos* e mais *cos*, que conflúen co plural dialectal dos substantivos *con* e *don*.

Unidade ortográfica	Alternativa 1	Alternativa 2
da	Verbo	Contracción de preposición e artigo
no	Adverbio	Contracción de preposición e artigo
cos	Substantivo	Contracción de preposición e artigo
dos	Substantivo	Contracción de preposición e artigo

¹⁴⁵ Inicialmente este foi o punto de partida: a elaboración de posibles combinacións de clíticos dependendo da forma verbal concreta e da súa transitividade. Non obstante, isto supoñía un gasto en tempo que non compensaba por varios motivos: a) a maior parte das combinacións non aparecerían; b) levaba cara á normativización e dificultaba o recoñecemento de texto real, e c) bastaba con identificar a combinación de forma verbal cunha listaxe simple de clíticos, cuxa combinación está restrinxida por regras lingüísticas.

3.3.2. Datos sobre a ambigüidade no corpus de adestramento

Creemos interesante compendiar nunha única epígrafe os datos relativos á ambigüidade que ao longo deste capítulo fomos achegando para os distintos subcorpus que compoñen o corpus de adestramento e ver que conclusións podemos sacar deles.

Os datos cuantitativos extractados son os seguintes:

	<i>Kernel</i>	<i>Economía</i>	<i>Xenérico</i>
Total de formas distintas	1921	19 868	22 927
Total de formas+etiquetas+lemas diferentes	2143	21 980	25 167
Formas con 1 etiqueta	1776	18 203	21 160
Formas con 2 etiquetas	109	1363	1467
Formas con 3 etiquetas	16	209	204
Formas con 4 etiquetas	7	64	59
Formas con 5 etiquetas	5	20	24
Formas con 6 etiquetas	4	4	5
Formas con 7 etiquetas		1	2
Formas con 8 etiquetas		2	1
Formas con 9 etiquetas	3	1	3
Formas con 10 etiquetas			1
Formas con 12 etiquetas		1	
Formas con 13 etiquetas	1		
Formas con 16 etiquetas			1
% de formas ambiguas	7,54%	8,38	7,70
Nº medio de etiquetas por forma	1,11	1,10	1,09

Tres son os aspectos máis notables que se deducen dos datos anteriores:

- 1) Arredor do 8% das formas do corpus presentan algún tipo de ambigüidade.
- 2) Segundo vai aumentando o número de etiquetas que recibe unha forma, decrece o número de formas, de tal xeito que pode falarse dunha relación proporcional á inversa entre número de etiquetas e número de formas.

3) Entre o 6 e o 7,5% da porcentaxe total de formas ambiguas corresponde a formas con dúas etiquetas, mentres que as unidades que presentan máis de seis etiquetas non chegan á media ducia.

Se agora prestamos atención aos datos de tipo cualitativo e recollemos as unidades que presentan máis de cinco etiquetas, atopámonos coa seguinte distribución:

	<i>Kernel</i>	<i>Economía</i>	<i>Xenérico</i>
Formas con 5 etiquetas	lle, nós, quen, súas, ti	15, 20, 200, 30, 40, 50, 80, catro, cen, cinco, dez, moral, nos, o, oito, presente, segundo, seis, sí, tal	26, 4, 50, 60, cada, cal, catro, cen, central, cinco, doce, fora, máis, nacionalistas, nada, nove, nós, oito, protagonista, segundo, seu, tanto, tres, un
Formas con 6 etiquetas	seu, seus, súa, te	100, calquera, me, tres	cantos, dez, lles, seis, tal
Formas con 7 etiquetas		lle	demais, si
Formas con 8 etiquetas		quen, si	nos
Formas con 9 etiquetas	me, nos, vos	qué	calquera, lle, me
Formas con 10 etiquetas			quen
Formas con 12 etiquetas		que	
Formas con 13 etiquetas	que		
Formas con 16 etiquetas			que

A característica común a todas estas unidades, partilladas por subcorpus, é a súa clasificación en máis dunha categoría e a pertinencia de atributos nalgunha delas, o que provoca un número maior de etiquetas distintas.

A pluralidade dos posesivos de terceira procede, sobre todo, do valor e do número do posuidor, en tanto que a das unidades *moral*, *presente*, *cal*, *central* e *nacionalistas* orixínase no sincretismo xenérico e na súa clasificación en varias categorías; a multiplicidade de *segundo* débese á súa inclusión nas categorías Adxectivo, Conxunción, Numeral, Preposición e Substantivo, mentres que a de *tanto* (Adverbio, Indefinido e Substantivo) e *un* (Artigo, Indefinido, Numeral e Substantivo), ademais de á variedade categorial, débese ao atributo valor: no Indefinido e no Adverbio para *tanto* e no Numeral para *un*.

Pola súa banda, os numerais, ben en letra ben en díxito, explican a súa aparición entre as formas máis ambiguas pola combinación dos atributos valor, xénero e número.

Finalmente, todas as demais presentes na táboa anterior son formas invariables moi frecuentes na lingua, cuxa característica máis destacada é a súa pertenza a categorías pechadas –Pronome (*me, nos, vos, lle...*), Indefinido (*demais, calquera...*), Conxunción (*que*), Relativo (*que, quen*) ou interrogativo (*que, quen*)– e o seu funcionamento nos paradigmas nos que se insiren mediante atributos aos que estas responden non no nivel da palabra senón coa súa inclusión nun contexto concreto.

3.4. A ferramenta de desambiguación manual

A ferramenta escollida para levar a cabo as tarefas de revisión da etiquetaxe automática, ou sexa, a desambiguación manual, é un editor de XML de software libre, en concreto o *XMLmind XML Editor Standard Edition V2.9*, o cal xa se utiliza para a incorporación dos documentos ao CORGA, e supón, polo tanto, familiaridade coa ferramenta. Este editor permite definir tanto o estilo que se quere utilizar como a estrutura interna que se desexa aplicar para distintos tipos de documentos. É, pois, a mesma aplicación, só que cun grao máis de complexidade nas súas funcionalidades, a cal, dependendo do documento que se quere editar, presenta a estrutura dos textos etiquetados ou prescinde de todo o relativo á anotación. Isto é posible grazas á existencia da folla de estilos e á potencialidade que ofrece o editor para describir a estrutura e sintaxe dos distintos tipos de documentos mediante *definición de tipo de documento* ou DTD.

As DTD sérvennos para, mediante etiquetas, describir os elementos que constitúen un documento, a orde en que van eses elementos dentro dun documento e, se é posible o aninamento de elementos, que elementos poden ir dentro doutros.

O empregar unha linguaxe de marcaxe XML garantiza que se mantén a consistencia entre os distintos documentos que utilizan a mesma DTD, ademais de facilitar que sexan validados e que se coñeza a estrutura dos elementos que trae consigo cada documento.

O desexo de acadar un sistema de desambiguación manual visualmente claro que facilitase o labor viuse obstaculizado no seu inicio pola existencia dos dous tipos de ambigüidade segmental que acabamos de ver. Non bastaba cunha organización en *forma, etiqueta e lema* para cada unidade léxica ou elemento gramatical. Había que darlles cabida ás distintas posibilidades de segmentación, tanto a nivel de palabra como

a nivel de secuencia de palabras. É así que, xunto á *unidade* de análise que acolle a forma ortográfica ou secuencia multipalabra que se vai analizar, xorden no noso sistema as etiquetas *alternativas*, *alternativa* e mais *constituínte*.

A primeira delas, *alternativas*, acolle no seu interior en sucesivas *alternativa* tantas posibilidades de segmentación como posúa unha unidade léxica dada; a segunda, *constituínte*, emprégase para dar conta dos distintos elementos gramaticais que conforman a *unidade*, podendo ser un só, se a unidade non é descompoñible, ou tantos como elementos gramaticais estean amalgamados nunha unidade determinada.

Todo o texto que contén un documento está organizado en parágrafos e logo estes en oracións. A partir de aquí entra en xogo a parte da DTD definida para a anotación, que é a que segue:

```
<!ELEMENT oración (expresión, análise?)>
<!ATTLIST oración distinto (%valores_distinto;) #IMPLIED>
<!ELEMENT expresión (%texto;|distinto)*>
<!ELEMENT análise (análise_unidade+)>
<!ELEMENT análise_unidade (unidade, alternativas)>
<!ELEMENT unidade (%texto;)*>
<!ELEMENT alternativas (alternativa+)>
<!ELEMENT alternativa (constituínte+)>
<!ATTLIST alternativa válido (si) #IMPLIED>
<!ELEMENT constituínte (forma, etq_lemma+)>
<!ELEMENT forma (%texto;)*>
<!ELEMENT etq_lemma (etiqueta, lema)>
<!ATTLIST etq_lemma válido (si) #IMPLIED>
<!ELEMENT etiqueta (#PCDATA)>
<!ELEMENT lema (%texto;)*>
```

En liñas xerais os símbolos pegados ás etiquetas indican:

Un ? pegado a unha etiqueta significa que a etiqueta é opcional.

Un + pegado a unha etiqueta significa que a etiqueta pode aparecer unha o máis veces.

Un * pegado a unha etiqueta significa que a etiqueta pode non aparecer, ou aparecer unha ou máis veces.

Un | paralelo a unha frecha significa un *or* co que hai anteriormente.

A etiqueta <texto> significa que na etiqueta que se está definindo vai texto.

Describindo a DTD liña a liña interpretamos que:

1. <oración> é unha etiqueta válida que contén unha etiqueta <expresión> e pode conter outra denominada <análise>. A opcionalidade do elemento <análise> indícanola a presenza do interrogante.
2. O elemento <oración> pode conter unha etiqueta *distinto*. Os valores para esta foron definidos xa antes, pois non afectan só á etiquetaxe, e son os seguintes: *outra_lingua*, *outro_período*, *non_normativo* e *descoñecido*. A liña da DTD na que se recolle esta información é

```
<!ENTITY % valores_distinto "outra_lingua|outro_período|non_normativo|descoñecido">
```
3. <expresión> é un elemento válido. Contén *texto*, ou a etiqueta *distinto*. O * vén determinado polo feito de que ese *texto* pode incluír, por exemplo, as etiquetas <subíndice> ou <superíndice> que foron explicadas nun punto anterior da DTD.
4. <análise> é un elemento correcto que conterá unha ou máis etiquetas <análise_unidade>.
5. O elemento <análise_unidade> contén respectivamente as etiquetas <unidade> e <alternativas>.
6. O elemento <unidade> contén *texto*.
7. O elemento <alternativas> contén un ou máis elementos <alternativa>.
8. A etiqueta <alternativa> contén un ou máis elementos denominados <constituínte>
9. O elemento <alternativa> pode levar asociado un atributo cuxo valor é *válido* (*si*).
10. O elemento <constituínte> contén a etiqueta <forma> e unha ou máis etiquetas <etq_lemma>
11. O elemento <forma> contén *texto*.
12. O elemento <etq_lemma> contén á súa vez os elementos <etiqueta> e mais <lema>
13. O elemento <etq_lemma> pode levar asociado un atributo cuxo valor é *válido* (*si*).
14. O elemento <etiqueta > contén datos de tipo alfanumérico.
15. O elemento <lema> contén *texto*.

A diferenza entre o que denominamos *texto* para os elementos <unidade>, <forma> e <lema> e datos de tipo alfanumérico para o elemento <etiqueta> reside en

que o *texto* pode incluír tamén subíndices ou superíndices, mentres que estes non teñen sentido no interior de <etiqueta>. Así, o símbolo correspondente á unidade de superficie Km², por exemplo, –clasificado en XIADA coa etiqueta Zs00 lexible como *Categoría periférica, símbolo, xénero non aplica, número non aplica*– represéntase por cuestións de codificación no corpus como “Km<superíndice>2</superíndice>”. Pois ben, esa codificación debe ser asumida e tratada tamén na etiquetaxe á hora de determinar a *unidade*, a *forma* de análise e mais o *lema* que lle corresponde, motivo polo que *texto* é definido como unha combinación de caracteres alfanuméricos que pode incluír outros elementos no seu interior, mais isto non se estende á <etiqueta>, onde só se inclúe un número limitado de letras e números, cuxa combinación motivada dá conta da caracterización morfolóxica concreta.

Recollemos a continuación un caso real, no que destacamos en negra o símbolo completo tal e como o codificamos e mais a etiqueta que se lle asigna, para que se perciba visualmente a distinción da que falamos:

```

<oración>
  <expresión>Estamos falando dun espazo económico de 51.000
Km<superíndice>2</superíndice> e 6,5 millóns de consumidores.</expresión>
</oración>
<análise>
  <análise_unidade>
    <unidade>Estamos</unidade>
    <constituínte>
      <forma>estamos</forma>
      <etiqueta>Vpi10p</etiqueta>
      <lema>estar</lema>
    </constituínte>
  </análise_unidade>
  <análise_unidade>
    [...]
  <análise_unidade>
    <unidade>Km<superíndice>2</superíndice></unidade>
    <constituínte>
      <forma>Km<superíndice>2</superíndice></forma>
      <etiqueta>Zs00</etiqueta>
      <lema>Km<superíndice>2</superíndice></lema>
    </constituínte>
  </análise_unidade>
</análise_unidade>
</análise_unidade>
</análise_unidade>
  [...]

```

A distinción é máis fácil de observar na edición do editor xml, pois á marxe de que cada un dos elementos que conforman a estrutura de análise definida está englobado en caixiñas que permiten ocultar as etiquetas de peche de cada un dos elementos, a cor

vermella, típica das etiquetas definidas na DTD, contrasta coa cor negra escollida para a compleción do contido desas mesmas etiquetas:

Oración ▾
 Expresión
 Estamos falando dun espazo económico de 51.000 Km<Superíndice>2</Superíndice> e 6,5 millóns de consumidores.

Análise ▾

Análise unidade ▾
 Unidade
 Estamos
 Constituínte ▾
 Forma estamos Etiqueta Vpi10p Lema estar

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾

Análise unidade ▾
 Unidade
 Km<Superíndice>2</Superíndice>
 Constituínte ▾
 Forma Km<Superíndice>2</Superíndice> Etiqueta Zs00 Lema Km<Superíndice>2</Superíndice>

Análise unidade ▾

Análise unidade ▾

Se agora trasladamos á práctica da etiquetaxe a lectura da DTD, comprenderemos mellor o funcionamento da estrutura e sintaxe xml definida. Comezaremos pola parte alta da DTD e logo descenderemos á análise das formas, onde irá aumentando progresivamente a complexidade a medida que vaiamos exemplificando.

Polo que respecta á unidade <oración> ofrecemos as seguintes posibilidades:

1) É unha secuencia que se desagrega en <expresión> e <análise>, contendo este último elemento tantas <análise_unidade> como unidades léxicas posúa a <expresión>:

```
<oración>
  <expresión>¿Que tiña eu de distinto para ser o escollido?</expresión>
  <análise>
    <análise_unidade>
      <unidade></unidade>
```

[...]

Oración ▾ Expresión ¿Que tiña eu de distinto para ser o escollido? Análise ▾
--

2) É unha secuencia que está nunha lingua distinta do galego e polo tanto non se analiza. A unidade <oración> contén un atributo denominado *distinto* caracterizado cun dos valores dunha lista pechada, neste caso *outra_lingua*:

```
<oración distinto="outra_lingua">
  <expresión>"¿Que cogiste ahí en la explanada?"</expresión>
</oración>
```

Oración ▾ Distinto: <input type="text" value="outra_lingua"/> ▾ Expresión "¿Que cogiste ahí en la explanada?"

3) A <expresión> contén un segmento textual caracterizado co atributo *distinto* ao lado de texto normal. A <análise> contén só as análises das unidades léxicas non comprendidas entre o segmento marcado con *distinto*:

```
<expresión><distinto tipo="outra_lingua">"Primavera, la espera</distinto> _cantaba un
vocalista con voz clara e cristalina_;</expresión>
<análise>
  <análise_unidade>
    <unidade>_</unidade>
  [...]
  </análise_unidade>
  <análise_unidade>
    <unidade>cantaba</unidade>
  [...]
</análise>
```

Oración ▾ Expresión Distinto Tipo: <input type="text" value="outra_lingua"/> ▾ "Primavera, la espera _cantaba un vocalista con voz clara e cristalina_; Análise ▾
Análise unidade ▾ Unidade _ Alternativas ▾ Alternativa ▾ Válido: <input checked="" type="checkbox"/> si
Análise unidade ▾ Unidade cantaba Alternativas ▾ Alternativa ▾ Válido: <input checked="" type="checkbox"/> si

4 Os atributos *inicio_poema* e *novo_verso* reciben o mesmo tratamento ca *distinto*: o tramo textual comprendido entre eles non se analiza, e tanto poden abranger a <expresión> completa como só un fragmento dela. Os dous seguintes exemplos mostran os dous subtipos:

```
<oración>
  <expresión><inicio_poema />No medio dun vento baleiro<novo_verso />que move os
  adeuses<novo_verso />un home agarda.</expresión>
</oración>
```

Oración ▾
 Expresión
 <Inicio poema>No medio dun vento baleiro<Novo verso>que move os adeuses<Novo verso>un home agarda.

```
<oración>
  <expresión><inicio_poema />alguén decidiu emprender unha viaxe<novo_verso />e
  converterse nun Vello camiñante<novo_verso />na procura das chaves do
  Labirinto<fin_poema />, e con ela as paisaxes dos lugares por onde pasa, a natureza e,
  en certa maneira, os homes.</expresión>
  <análise>
    <análise_unidade>
      <unidade>,</unidade>
  [...]
    <análise_unidade>
      <unidade>e</unidade>
  [...]
    <análise_unidade>
      <unidade>ela</unidade>
  [...]
```

Oración ▾
 Expresión
 <Inicio poema>alguén decidiu emprender unha viaxe<Novo verso>e converterse nun Vello camiñante<Novo verso>na procura das chaves do Labirinto<Fin poema>, e con ela as paisaxes dos lugares por onde pasa, a natureza e, en certa maneira, os homes.

Análise ▾

Análise unidade ▾
 Unidade
 ,
 Constituinte ▾
 Forma , Etiqueta Q, Lema ,

Análise unidade ▾
 Unidade
 e
 Constituinte ▾
 Forma e Etiqueta Cc Lema e

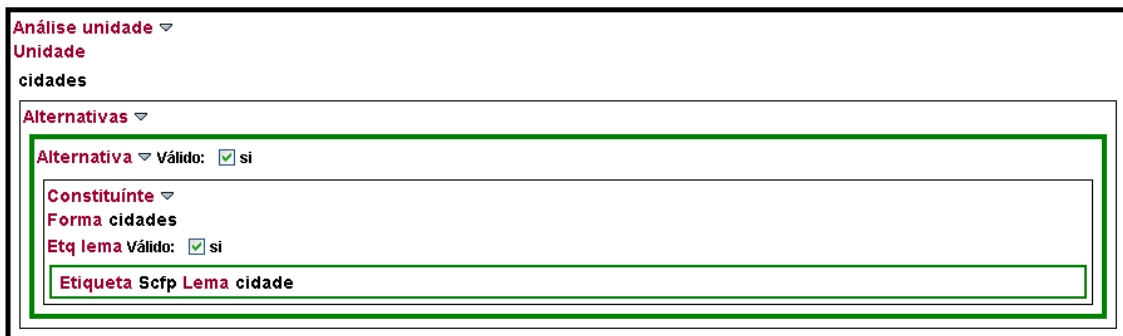
En relación cos elementos <unidade>, encontrámonos con posibilidades estruturais con diverso grao de complexidade. Son as seguintes:

1) A unidade léxica non presenta ambigüidade de ningún tipo. Nela non hai ambigüidade segmental, polo que só se achega unha <alternativa>; non é

descompoñible en elementos gramaticais, polo que só presenta un <constituínte> e, finalmente, non hai ambigüidade categorial ou atributiva polo que só pode caracterizarse cunha <etiqueta> e remitirse a un único lema. Así, a unidade *ciudades* caracterízase coa <etiqueta> Scfp (*Substantivo, común, feminino, plural*) e remítese ao lema *cidade*:

```
<análise_unidade>
  <unidade>ciudades</unidade>
  <alternativas>
    <alternativa válido="si">
      <constituínte>
        <forma>ciudades</forma>
        <etq_ lema válido="si">
          <etiqueta>Scfp</etiqueta>
          <lema>cidade</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
  </alternativas>
</análise_unidade>
```

A vista no editor xml é moito máis amable co usuario:



Outros exemplos poderían ser as unidades léxicas *tipo*, *candil*, *perviviu*, *en*, *hai*, *demográfico*, etc.

2) A unidade léxica presenta ambigüidade categorial, polo que a súa análise comprenderá tantos elementos <etq_ lema>, cos correspondentes <etiqueta> e <lema>, como categorías diferentes ás que pode responder, salvo que só un deles levará o atributo *válido*, o cal dá por boa esa análise. O exemplo seguinte corresponde á unidade *chegada* no segmento *A chegada da luz*, a cal pode ser analizada, dependendo do contexto, como

i) A0fs (*Adxectivo, grao non aplica, femino, singular*) cuxo lema é *chegado*;

ii) V0p0fs (*Verbo, tempo non aplica, participio, persoa non aplica, feminino, singular*) cuxo lema é *chegar*;

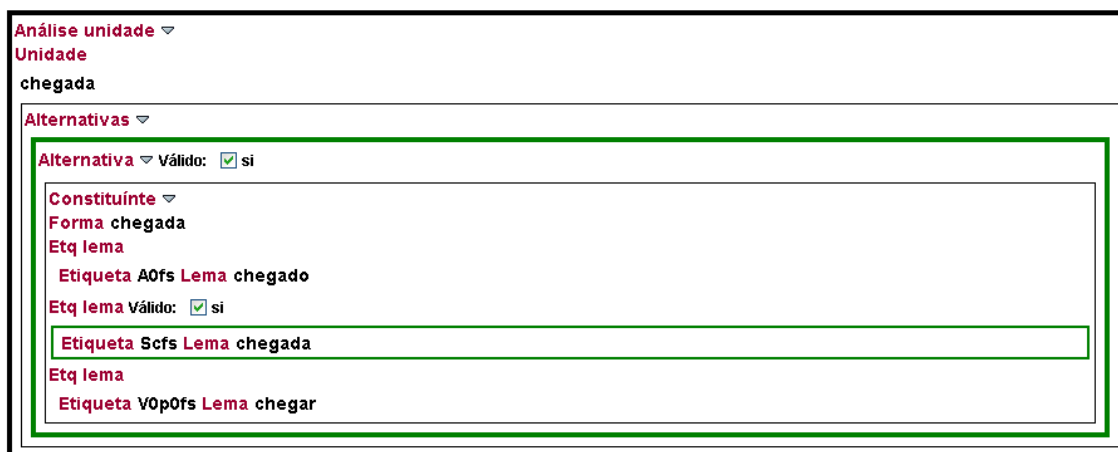
iii) a correcta no contexto de onde a extraemos, Scfs (*Substantivo, común, feminino, singular*) cuxo lema é *chegada*.

```

<análise_unidade>
  <unidade>chegada</unidade>
  <alternativas>
    <alternativa válido="si">
      <constituínte>
        <forma>chegada</forma>
        <etq_ lema>
          <etiqueta>A0fs</etiqueta>
          <lema>chegado</lema>
        </etq_ lema>
        <etq_ lema válido="si">
          <etiqueta>Scfs</etiqueta>
          <lema>chegada</lema>
        </etq_ lema>
        <etq_ lema>
          <etiqueta>V0p0fs</etiqueta>
          <lema>chegar</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
  </alternativas>
</análise_unidade>

```

A visualización no editor, por suposto, é máis amigable:



Outros exemplos son os elementos gramaticais *esquerda, cera, figuras, alto, a, case, galegas*, etc.

3) A unidade presenta ambigüidade atributiva, polo que a súa análise comprenderá tantos elementos <etq_ lema>, cos correspondentes <etiqueta> e <lema>, como etiquetas diferentes ás que poida responder, mais só un deles levará o atributo

válido cuxo valor é *si*. O exemplo seguinte corresponde á unidade léxica *meirande* inserida no segmento *na meirande parte*, a que, en función do contexto, pode ser caracterizada como adxectivo en grao comparativo –Ac– feminino singular (Acfs), masculino singular (Acms) ou ambiguo en canto ao xénero (Acas):

```

<análise_unidade>
  <unidade>meirande</unidade>
  <alternativas>
    <alternativa válido="si">
      <constituínte>
        <forma>meirande</forma>
        <etq_ lema>
          <etiqueta>Acas</etiqueta>
          <lema>meirande</lema>
        </etq_ lema>
        <etq_ lema válido="si">
          <etiqueta>Acfs</etiqueta>
          <lema>meirande</lema>
        </etq_ lema>
        <etq_ lema>
          <etiqueta>Acms</etiqueta>
          <lema>meirande</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
  </alternativas>
</análise_unidade>

```

The screenshot displays a hierarchical view of the analysis for the unit 'meirande'. At the top, 'Análise unidade' is expanded to show 'Unidade meirande'. Below this, 'Alternativas' is expanded to show three alternatives, each with a 'Válido' checkbox checked. The first alternative is 'Alternativa Válido: si', which is expanded to show 'Constituínte' and 'Forma meirande'. Under 'Constituínte', there are three 'Etq lema' entries: 'Etiqueta Acas Lema meirande', 'Etiqueta Acfs Lema meirande', and 'Etiqueta Acms Lema meirande'. Each of these 'Etq lema' entries has a 'Válido: si' checkbox checked.

4) A unidade está formada por máis dun elemento gramatical. Cada un deles aparece identificado coa etiqueta <constituínte>. Esta análise corresponde coa das contraccións e formas verbais con pronomes enclíticos e ou artigo non ambiguas; é dicir, só hai unha posibilidade de análise para cada un dos elementos que compoñen a amalgama. A particularidade deste subtipo é a segregación en <constituínte> dos formantes da unidade amalgamada. Por exemplo, a análise da contracción *das* ou da

forma verbal con enclítico *pódese* recolle a segmentación en dous constituíntes con análises inequívocas:

das: preposición *de* (P) con lema *de* e mais o artigo *as* (Ddfp: *Artigo, determinado, feminino plural*) cuxo lema é *o*:

```
<análise_unidade>
<unidade>das</unidade>
<alternativas>
<alternativa válido="si">
<constituínte>
<forma>de</forma>
<etq_ lema válido="si">
<etiqueta>P</etiqueta>
<lema>de</lema>
</etq_ lema>
</constituínte>
<constituínte>
<forma>as</forma>
<etq_ lema válido="si">
<etiqueta>Ddfp</etiqueta>
<lema>o</lema>
</etq_ lema>
</constituínte>
</alternativa>
</alternativas>
</análise_unidade>
```

pódese: terceira persoa do singular do presente de indicativo (Vpi30s) de *poder* e mais o pronome *se*, ao cal lle aplicamos en XIADA unha caracterización única (Rao3aa: *Pronome, átono, outros, terceira, feminino/masculino, singular/plural*):

```
<análise_unidade>
<unidade>pódese</unidade>
<alternativas>
<alternativa válido="si">
<constituínte>
<forma>pode</forma>
<etq_ lema válido="si">
<etiqueta>Vpi30s</etiqueta>
<lema>poder</lema>
</etq_ lema>
</constituínte>
<constituínte>
<forma>se</forma>
<etq_ lema válido="si">
<etiqueta>Rao3aa</etiqueta>
<lema>se</lema>
</etq_ lema>
</constituínte>
</alternativa>
</alternativas>
```


</análise_unidade>

As análises respectivas a este subgrupo visualízanse no editor xml como segue:

Análise unidade ▾
Unidade
das

Alternativas ▾

Alternativa ▾ Válido: si

Constituinte ▾
Forma de
Etq lema Válido: si
Etiqueta P Lema de

Constituinte ▾
Forma as
Etq lema Válido: si
Etiqueta Ddfp Lema o

Análise unidade ▾
Unidade
pódese

Alternativas ▾

Alternativa ▾ Válido: si

Constituinte ▾
Forma pode
Etq lema Válido: si
Etiqueta Vpi30s Lema poder

Constituinte ▾
Forma se
Etq lema Válido: si
Etiqueta Rao3aa Lema se

5) A unidade léxica presenta ambigüidade segmental e, en función do contexto, pode corresponder unha única unidade ou pode ser segmentable en máis dun elemento lingüístico. Así, a análise da unidade contén, como mínimo, dúas alternativas, recollidas en cadansúa <alternativa>. Mostras destoutro subtipo son *á*, *coas*, *falo*, *vela*, *tela*, etc. Así, a unidade *coas* correspóndese co elemento gramatical *coas* cuxa análise é a de segunda persoa do presente de indicativo do verbo *coar* ou correspóndese coa análise dun primeiro elemento gramatical *con*, que é preposición, e mais dun segundo *as*, que é o artigo determinado feminino plural cuxo lema é *o*:

```
<análise_unidade>
  <unidade>coas</unidade>
  <alternativas>
    <alternativa>
      <constituínte>
```

```

<forma>coas</forma>
<etq_lemma>
  <etiqueta>Vpi20s</etiqueta>
  <lema>coar</lema>
</etq_lemma>
</constituente>
</alternativa>
<alternativa válido="si">
  <constituente>
    <forma>con</forma>
    <etq_lemma válido="si">
      <etiqueta>P</etiqueta>
      <lema>con</lema>
    </etq_lemma>
  </constituente>
  <constituente>
    <forma>as</forma>
    <etq_lemma válido="si">
      <etiqueta>Ddfp</etiqueta>
      <lema>o</lema>
    </etq_lemma>
  </constituente>
</alternativa>
</alternativas>
</análise_unidade>

```

No editor xml vémola do seguinte xeito:

Análise unidade ▾
 Unidade
 coas

Alternativas ▾

Alternativa ▾

Constituente ▾
 Forma coas
 Etq lema
 Etiqueta Vpi20s Lema coar

Alternativa ▾ Válido: si

Constituente ▾
 Forma con
 Etq lema Válido: si
 Etiqueta P Lema con

Constituente ▾
 Forma as
 Etq lema Válido: si
 Etiqueta Ddfp Lema o

6) A unidade é ambigua segmentalmente, pero a maiores, a diferenza do grupo anterior, un dos constituíntes pode presentar máis dun <etq_lemma>, motivado por integrarse o elemento gramatical en máis dunha categoría ou por variar algún dos

atributos da súa etiqueta. Dáse, por exemplo, coas contraccións nas que participa *un* ou con *leo*, *coa*, *nos*, etc.

Se tomamos como mostra a unidade *leo*, esta pode ser tratada como unha soa unidade léxica: *leo*, caracterizable como primeira persoa singular do presente de indicativo (Vpi10s) dos verbos *lear* ou *ler*; mais tamén pode corresponder á amálgama da forma verbal *le* e mais o pronome de acusativo de terceira masculino singular, sendo ademais *le* caracterizable como terceira persoa singular do presente de indicativo (Vpi30s) ou como segunda persoa singular do imperativo (V0m20s) do verbo *ler*:

```

<análise_unidade>
  <unidade>leo</unidade>
  <alternativas>
    <alternativa>
      <constituínte>
        <forma>leo</forma>
        <etq_ lema>
          <etiqueta>Vpi10s</etiqueta>
          <lema>lear</lema>
        </etq_ lema>
        <etq_ lema>
          <etiqueta>Vpi10s</etiqueta>
          <lema>ler</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
    <alternativa válido="si">
      <constituínte>
        <forma>le</forma>
        <etq_ lema>
          <etiqueta>V0m20s</etiqueta>
          <lema>ler</lema>
        </etq_ lema>
        <etq_ lema válido="si">
          <etiqueta>Vpi30s</etiqueta>
          <lema>ler</lema>
        </etq_ lema>
      </constituínte>
      <constituínte>
        <forma>o</forma>
        <etq_ lema válido="si">
          <etiqueta>Raa3ms</etiqueta>
          <lema>o</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
  </alternativas>
</análise_unidade>

```

Novamente, a visualización é moito máis cómoda a través do editor xml:

Análise unidade ▾
Unidade
leo

Alternativas ▾

Alternativa ▾

Constituínte ▾
Forma leo
Etq lema
Etiqueta Vpi10s Lema ler
Etq lema
Etiqueta Vpi10s Lema lear

Alternativa ▾ Válido: si

Constituínte ▾
Forma le
Etq lema
Etiqueta V0m20s Lema ler
Etq lema Válido: si
Etiqueta Vpi30s Lema ler

Constituínte ▾
Forma o
Etq lema Válido: si
Etiqueta Raa3ms Lema o

7) Finalmente queda o esquema daquelas unidades que presentan ambigüidade segmental no nivel superior á palabra. Falamos das que, en función do contexto, poden analizarse como unha unidade multipalabra ou ser descompoñibles en tantas unidades léxicas como unidades ortográficas presente o segmento. Neste grupo entran todas as locucións caracterizadas como inseguras no lexicón, mentres que as seguras van ao primeiro dos esquemas descritos.

Este esquema presenta unha incongruencia insalvable: aparecen baixo <constituínte> elementos gramaticais que en realidade non o son, pois non estamos ante unha amálgama. Non obstante, dado que é imposible dar cabida ao tempo á análise multipalabra e ás análises individuais, en XIADA adoptamos unha solución de compromiso: a análise pertencente ao tratamento como unidades independentes aparece baixo unha <alternativa> que posúe tantos constituíntes como unidades léxicas presente o segmento potencialmente multipalabra.

Esta actuación, porén, non repercute no etiquetador nin no seu adestramento, posto que este non ten en conta as formas ortográficas recollidas nos elementos <unidade> e si as unidades léxicas ou elementos gramaticais recollidos nos elementos <forma>. Redúcese, pois, a un problema meramente teórico e de representación que non

ten máis consecuencias ca a visualización en distintos constituíntes de formas léxicas que deberían presentar tantas <análise_unidade> como elementos diferentes existan.

Un exemplo do esquema que tratamos constitúeo a análise de *por certo*. Só o contexto nos vai permitir diferenciar entre as seguintes posibilidades de análise:

<alternativa> 1: locución adverbial que actúa como marcador de tipo discursivo, recollida en primeiro lugar e marcada como válida;

<alternativa> 2:

2.a) preposición seguida do indefinido con valor *determinante*, ou sexa seguido por un substantivo;

2.b) preposición seguida do adxectivo *certo*, como sinónimo de *verdadeiro*;

2.c) preposición seguida do adverbio *certo*¹⁴⁶;

```

<análise_unidade>
  <unidade>por certo</unidade>
  <alternativas>
    <alternativa válido="si">
      <constituínte>
        <forma>por certo</forma>
        <etq_ lema válido="si">
          <etiqueta>La0</etiqueta>
          <lema>por certo</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
    <alternativa>
      <constituínte>
        <forma>por</forma>
        <etq_ lema válido="si">
          <etiqueta>P</etiqueta>
          <lema>por</lema>
        </etq_ lema>
      </constituínte>
      <constituínte>
        <forma>certo</forma>
        <etq_ lema>
          <etiqueta>A0ms</etiqueta>
          <lema>certo</lema>
        </etq_ lema>
        <etq_ lema>
          <etiqueta>Idms</etiqueta>
          <lema>certo</lema>
        </etq_ lema>
        <etq_ lema>
          <etiqueta>Wn</etiqueta>

```

¹⁴⁶ Esta possibilidade de análise para *certo*, a de adverbio, non é factible neste contexto; non obstante, aparece porque, como xa sabemos, o etiquetador ofrécenos todas as opcións de análise para cada forma.

```

    <lema>certo</lema>
  </etq_ lema>
</constituínte>
</alternativa>
</alternativas>
</análise_unidade>

```

A diferenza coas unidades multipalabra seguras é enorme. Vémolo por exemplo na análise de *a esgalla*, onde só hai unha alternativa e un constituínte cos seus respectivos elementos <forma>, <etiqueta> e <lema>

```

<análise_unidade>
  <unidade>a esgalla</unidade>
  <alternativas>
    <alternativa válido="si">
      <constituínte>
        <forma>a esgalla</forma>
        <etq_ lema válido="si">
          <etiqueta>La0</etiqueta>
          <lema>a esgalla</lema>
        </etq_ lema>
      </constituínte>
    </alternativa>
  </alternativas>
</análise_unidade>

```

A diferenza entre as dúas análises, a da locución segura e a da insegura, malia que máis fácil de percibir visualmente, tamén se observa con claridade no editor xml:

Análise unidade ▾
 Unidade
 por certo

Alternativas ▾

Alternativa ▾ Válido: si

Constituínte ▾
 Forma por certo
 Etq lema válido: si
 Etiqueta La0 Lema por certo

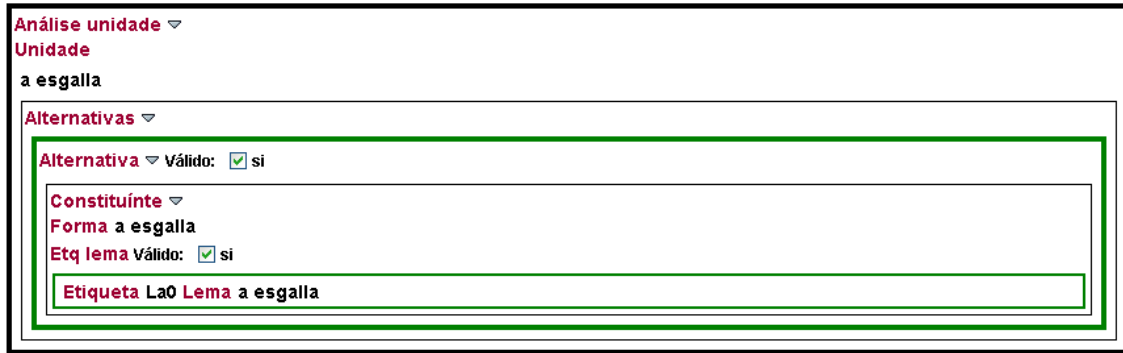
Alternativa ▾

Constituínte ▾
 Forma por
 Etq lema
 Etiqueta P Lema por

Constituínte ▾
 Forma certo
 Etq lema
 Etiqueta A0ms Lema certo

Etq lema
 Etiqueta Idms Lema certo

Etq lema
 Etiqueta Wn Lema certo



Un exemplo integrado neste mesmo esquema pero que polo contexto non corresponde á análise multipalabra, senón a unidades léxicas contiguas, é a de *de todo* no segmento *acompañados de todo tipo de ruidos*. Aquí, *de todo* non pode considerarse locución adverbial parafraseable por *por completo*, senón que cada unidade ortográfica debe ser considerada unha unidade léxica. Na representación desta análise observamos a incongruencia visual, que non real, de tratar como <constituínte> a *de* e *todo*, tanto no formato de só texto como no editor xml:

```

<análise_unidade>
  <unidade>de todo</unidade>
  <alternativas>
    <alternativa>
      <constituínte>
        <forma>de todo</forma>
        <etq_lema>
          <etiqueta>La0</etiqueta>
          <lema>de todo</lema>
        </etq_lema>
      </constituínte>
    </alternativa>
    <alternativa válido="si">
      <constituínte>
        <forma>de</forma>
        <etq_lema válido="si">
          <etiqueta>P</etiqueta>
          <lema>de</lema>
        </etq_lema>
        <etq_lema>
          <etiqueta>Scms</etiqueta>
          <lema>de</lema>
        </etq_lema>
      </constituínte>
      <constituínte>
        <forma>todo</forma>
        <etq_lema>
          <etiqueta>Iams</etiqueta>
          <lema>todo</lema>
        </etq_lema>
        <etq_lema válido="si">
          <etiqueta>Idms</etiqueta>

```

```

<lema>todo</lema>
</etq_lema>
<etq_lema>
<etiqueta>Inms</etiqueta>
<lema>todo</lema>
</etq_lema>
</constituínte>
</alternativa>
</alternativas>
</análise_unidade>

```

Análise unidade ▾

Unidade

de todo

Alternativas ▾

Alternativa ▾

Constituínte ▾

Forma de todo

Etq lema

Etiqueta La0 Lema de todo

Alternativa ▾ **Válido:** **si**

Constituínte ▾

Forma de

Etq lema **Válido:** **si**

Etiqueta P Lema de

Etq lema

Etiqueta Scms Lema de

Constituínte ▾

Forma todo

Etq lema

Etiqueta lams Lema todo

Etq lema **Válido:** **si**

Etiqueta Idms Lema todo

Etq lema

Etiqueta Inms Lema todo

En realidade, algo semellante ocorre coas locucións preposicionais nas que está implicada unha contracción, só que neste caso, precisamente pola amálgama, non nos sorprende ver dous elementos gramaticais na mesma unidade:

```

<análise_unidade>
<unidade>a pesar das</unidade>
<alternativas>
<alternativa válido="si">
<constituínte>
<forma>a pesar de</forma>
<etq_lema válido="si">
<etiqueta>Lp0</etiqueta>
<lema>a pesar de</lema>
</etq_lema>
</constituínte>
<constituínte>

```



```

<forma>as</forma>
<etq_lemma válido="si">
  <etiqueta>Ddfp</etiqueta>
  <lemma>o</lemma>
</etq_lemma>
</constituente>
</alternativa>
</alternativas>
</análise_unidade>

```

Análise unidade ▾
 Unidade
 a pesar das

Alternativas ▾

Alternativa ▾ Válido: si

Constituente ▾
 Forma a pesar de
 Etq lemma Válido: si
 Etiqueta LpO Lema a pesar de

Constituente ▾
 Forma as
 Etq lemma Válido: si
 Etiqueta Ddfp Lema o

Por outra banda, ademais das funcionalidades de calquera editor, o *XMLmind* permítenos comprobar en todo momento a adecuación do documento á DTD mediante a opción de validado. Así mesmo, a creación de novos elementos `<unidade>`, `<constituente>`, `<etiqueta>`, etc. só é posible no nivel de análise correcto, pois tamén está controlada pola DTD, mentres que a supresión ou modificación incorrecta do contido deses mesmos elementos, malia poderse realizar tamén dende a DTD, preferimos supervisala dende o exterior a través de scripts, tarefa da que se ocupa a parte informática do proxecto, que controlan, sobre todo, que

- a etiqueta asignada sexa unha das que se orixinan no etiquetario. Esta comprobación fai que salten aquelas etiquetas que por erro humano, ao modificalas ou introducilas manualmente, levan algún carácter de máis ou de menos, mais tamén pode ocorrer que a etiqueta sexa correcta e haxa que darlle cabida no etiquetario.
- figure a información relativa a `<forma>` e `<lemma>`. Non sería estraño que ao manipular a segmentación dalgunha unidade se esquecese introducir a `<forma>` ou

completar o <lema> en casos de palabras descoñecidas ou de análises que non corresponden á realizada polo etiquetador.

- en cada <unidade> estean marcados co atributo *válido si* unha <alternativa> e mais un <etq_lema>. Na saída do documento etiquetado automaticamente sempre hai unha <alternativa> e un <etq_lema> marcados como válidos, non obstante, cando a solución escollida polo etiquetador non é a correcta e polo tanto hai que modificar a análise da unidade para corríxila, pode pasar que nos esquezamos de marcar como válida a alternativa correcta ou a etiqueta e lema que lle corresponden, ou ambos, ou que se manteña a validez para máis dunha alternativa e/ou o par etiqueta/lema. O script detecta calquera das variantes incorrectas e sinálaas.

No CD que acompaña este traballo inclúese unha noticia do semanario *A Nosa Terra* en formato XML, xunto cunha copia do *XMLmind XML Editor Standard Edition V2.9*, para que poidan observarse tanto as características do propio editor como as diferenzas entre as tres fases polas que transcorre a etiquetaxe dun documento cando hai supervisión manual:

- 1) arquivo etiquetado automaticamente: ANT2008_07_10_57_orixinal.etq.xml;
- 2) arquivo revisado manualmente: ANT2008_07_10_57_desambiguado.etq.xml;
- e 3) o documento etiquetado final: ANT2008_07_10_57_definitivo.etq.xml.

3.5. Conclusións

Os principais puntos que podemos destacar tras a lectura deste capítulo son os seguintes:

1) Os sistemas de etiquetaxe fundamentados no método estatístico revélanse como os máis axeitados para o procesamento da linguaxe natural debido ás facilidades de adestramento e ao bo rendemento. A pecha de situar o seu límite de acerto no 97% sálvase coa construción de regras lingüísticas específicas que corríxen erros de etiquetaxe recorrentes cos que a probabilística non atina.

2) É preciso que, para obter bos resultados, o etiquetador conte cun corpus de adestramento amplo e variado acorde lingüisticamente cos textos que se vaian etiquetar. Este é un factor cuxa relevancia se manifesta superior á da do tamaño do etiquetario.

3) Ao redor do 8% das unidades léxicas do corpus de adestramento son ambiguas. As unidades léxicas cun maior número de etiquetas distintas posibles son *que, qué* e *quen*. As demais están tamén moi localizadas e relaciónanse con elementos de categorías pechadas: Indefinido (*de mais, tal, calquera...*), Pronome (*me, te, nos, nós, vós...*), Numeral (*dez, tres, seis...*), etc. Polo que respecta ás unidades ortográficas máis frecuentes con ambigüidade segmental a nivel de palabra, as análises implicadas son maioritariamente substantivos, adxectivos ou formas verbais que coinciden formalmente con algunha contracción ou coa amalgama de forma verbal e mais clítico; mentres que a nivel de segmento multipalabra, fronte ás análises individuais, figuran os diversos tipos de locución e mais os nomes propios.

4) Finalmente, a escolla dunha ferramenta de desambiguación que ofrece unhas funcionalidades de representación e tratamento dos datos tales que garanten que non van existir etiquetas imposibles nin contido baleiro de elementos incrementa a consistencia do corpus de adestramento, facilita o labor manual e permite *a posteriori* a validación automática dos datos finais.

Capítulo 4

A etiquetaxe manual

4.1. *Introdución*

Como vimos de ver no capítulo anterior, un etiquetador estatístico precisa dispoñer dun modelo de referencia a partir do cal obter os datos relativos ás unidades léxicas ou *tokens* e ás probabilidades de uso das diferentes etiquetas. Este modelo denomínase xeralmente en lingüística de corpus *gold standard* ou corpus de adestramento.

A razón de ser do corpus de adestramento fai que, precisamente para poder servir de modelo, este deba reunir unha serie de requisitos entre os que destacan a corrección e a coherencia da etiquetaxe realizada. Naturalmente, isto tamén provoca que nalgún momento do proceso apareza o ser bípede por excelencia para supervisar ou realizar por completo as tarefas de etiquetaxe, o que á súa vez favoreceu nos últimos anos o desenvolvemento de recursos e medidas destinadas a avaliar e mais minimizar o impacto das intervencións humanas erróneas¹⁴⁷.

O máis usual é que os corpus de adestramento sexan primeiro etiquetados automaticamente e logo revisados e corrixidos dun xeito manual ou semimanual por lingüistas; non obstante, tamén existe algún caso, como o corpus SUSANNE¹⁴⁸, no que a etiquetación se produce case na súa totalidade de forma manual (Graña, 2000: 25).

Con frecuencia, dado o tempo que requiren os labores de desambiguación, é un grupo de lingüistas quen os leva a cabo; no noso caso decidimos que fose unha soa persoa a que revisase manualmente todo o corpus de adestramento, pois deste xeito elimínanse as discrepancias entre anotadores, sistematízanse os posibles erros e increméntase a congruencia na etiquetaxe final, co que aumenta a calidade do corpus,

¹⁴⁷ Entre outros traballos, destacan os estudos sobre a consistencia na etiquetaxe morfolóxica de Sampson e Babarczy (2003) e mais Babarczy, Carroll e Sampson (2006) ou, xa no terreo da etiquetaxe sintáctica, o do grupo encabezado por Civit para o corpus Cast3LB (2003b), así como o de Sampson e Babarczy (2008).

¹⁴⁸ Véxase <http://www.grsampson.net/Resources.html> [consultado o 14/01/2011].

así como a súa utilidade tanto para o adestramento e test de sistemas de análise automática como para a investigación lingüística.

Posto que os sistemas de desambiguación baseados na aprendizaxe automática necesitan corpus previamente etiquetados para inferir o modelo de lingua que logo se empregará para desambiguar outros textos, e máis cando o propio corpus de adestramento se pon a disposición pública, é sumamente interesante coñecer os criterios que se seguiron na etiquetaxe manual.

A complexidade da problemática do etiquetado de corpus procede de varias fontes. Por unha parte, unha mesma estrutura lingüística é susceptible de diferentes interpretacións, todas elas correctas. Por outra, moitas das expresións lingüísticas que aparecen nos corpus, reflexo do uso da lingua, crean problemas que en numerosas ocasións non se ven reflectidos nas gramáticas ou son tratados de xeito parcial.

Neste capítulo imos tratar os criterios xerais de lematización e etiquetaxe, mais tamén os dalgúns casos especiais como son as datas, as horas, os números de teléfono, os enderezos web, etc. Así mesmo estableceremos os límites categoriais de *tokens* problemáticos como *un, unha, uns, que, como*, etc., sen deixar de lado as características dos textos escritos en galego e a cuestión dos erros gráficos, comezando polas limitacións que nos impón o propio sistema.

4.2. O contexto na desambiguación manual

A utilización dun etiquetador basicamente estatístico establece *a priori* un condicionante primordial sobre os resultados da etiquetación que se vai practicar e, en consecuencia, tamén inflúe na elaboración do protocolo que vai guiar a desambiguación manual. Trátase da esixencia de limitar o contexto que serve de marco da etiquetación. En xeral, a información que emprega un etiquetador estatístico para desambiguar unha palabra dada é o contexto, entendendo por este as unidades e etiquetas anteriores e/ou posteriores a ela. A maioría dos sistemas automáticos traballan con combinacións de dous ou tres pares de elementos unidade-etiqueta, os denominados respectivamente bigramas e trigramas.

Dado que nós non estamos seguros do alcance do etiquetador que usamos, decidimos establecer como límite contextual a oración, pois do que si estamos certos é

de que non ten en conta un contexto superior a ese. É por isto que nas directrices xerais do manual de desambiguación establecemos que naquelas clases de palabras nas cales un determinado atributo é pertinente, malia que nalgunhas das unidades nelas clasificables non se manifeste un valor claro a nivel de palabra, atribuíraselle o valor individual clasificatorio sempre que algún elemento do contexto correspondente a esa unidade, é dicir explícito na oración, nos proporcione a información necesaria para concretar o valor, sexa este de xénero, número, persoa, etc.

Esta consigna teórica aplicarémola non só aos substantivos e adxectivos con xénero común ou invariables, senón tamén en categorías coma Pronome, Relativo, Indefinido, etc., nas que é inusual clasificar os membros do paradigma que non presentan formalmente flexión de xénero ou número.

Deste xeito, por exemplo, para establecer na etiquetaxe de corpus o xénero dos substantivos comúns que presentan, nas obras lexicográficas na súa caracterización morfolóxica, unha soa forma para os dous xéneros gramaticais terase en conta a súa aparición en cada enunciado concreto para, por medio dos determinantes ou adxectivos que presentan variación xenérica, determinar o xénero do substantivo que se corresponde co sexo do referente. Nos seguintes enunciados *xornalista* é caracterizado como feminino na primeira oración fronte ao masculino da segunda, mentres que na última tanto a *socialdemócrata* como a *galeguista* se lles asigna o atributo xenérico ambiguo masculino/feminino.

Na mesma liña, a xornalista destacou que o programa "mantívose fiel" a uns criterios, que lle permitiron manter esa fidelidade, esa credibilidade e ese agarimo aos seus contidos. [CORGAetq: GH2008-08-05/102]

Admite que ante o lector, o xornalista está perdendo credibilidade, algo que "está sucedendo na prensa, porque os clientes non tragan o aliñamento político". [CORGAetq: GH2007-03-30/92]

Nos meus tempos no Partido Nacionalista Galego sempre me reclamaba socialdemócrata e galeguista. [CORGAetq: ANT2005-03-03/19]

Outro exemplo do atributo xénero, este pertencente á clase Pronome, amósanos a aplicación deste criterio en elementos que normalmente fican sen especificar, malia que os condicionantes para o seu establecemento son practicamente os mesmos que operan nos substantivos ou adxectivos. Así, o xénero do pronome tónico *eu* individualízase no seu valor masculino ou feminino se na oración figura algún elemento que nos

permita identificalo cun referente masculino ou feminino; en caso contrario permanecerá sen concretar co valor ambiguo masculino/feminino. Vexámolo aplicado na etiquetación do CORGAetq. Nas tres oracións que seguen aparece o pronome *eu*, mais o valor do atributo xénero diverxe en todas elas. No primeiro caso, o substantivo feminino *mestra* acláranos que o referente pertence ao sexo feminino, polo que o caracterizamos como feminino; no segundo, o adxectivo masculino sinálanos o sexo do referente do pronome *e*, no terceiro, a ausencia de elementos que aclaren o xénero dese *eu* provoca que o determinemos co valor ambiguo masculino/feminino:

Pola miña parte, eu son mestra en ikebana polas escolas Sogetsu e Koriu, despois de sete anos de práctica no Xapón. [CORGAetq: ANT2005-03-03/52]

"Estou convencido de que non aceptar esta oferta é un erro que conducirá á compañía a un lugar a onde eu non quero levala", sinalou Hilbers. [CORGAetq: CG2002-05-19/55]

El e mais eu non nos parecemos en nada. [CORGAetq: DLV2007-05-15/108]

4.3. Características dos textos: erratas, variacións, dialectalismos, hipergaleguismos, castelanismos, etc.

Unha das decisións máis complexas que se deben tomar cando se constrúe ou etiqueta un corpus é como actuar ante as distintas desviacións da norma e os erros que se rexistran nos documentos orixenarios.

En xeral, é inevitable no ser humano, cando este escribe, cometer distintas transgresións contra a lingua. Por moito que revisemos o texto escrito, sempre se nos escapará algunha gralla ou nos equivocaremos nalgún aspecto morfosintáctico. Se ademais imos tratar, como é o noso caso, con textos escritos nunha lingua que aínda está en proceso de normalización os problemas acrecéntanse. Por riba, aos condicionantes anteriores debemos engadir-lles, aínda, a premura de tempo característica dos medios xornalísticos¹⁴⁹ e, nos últimos anos, a existencia de correctores e tradutores automáticos, cuxa combinación provoca a aparición nos textos de formas realmente estrañas para a nosa lingua.

A decisión que tomamos cando comezamos a construír o CORGA determinamos mantela no corpus de adestramento: non manipularemos os documentos. Etiquetaremos

¹⁴⁹ O protocolo de desambiguación elabórase tomando como punto de partida os textos xornalísticos, dado que ese estilo conforma o primeiro paquete de adestramento do noso etiquetador.

o que neles se nos presente, sen ningún tipo de marcas, independentemente de que as unidades sexan normativas ou non, aparezan no VOLGa ou no DRAG ou non. O noso obxectivo é etiquetar un corpus real descritivo da lingua galega actual e non inventar ou construír un corpus ideal. Así, as variacións dialectais, os castelanismos, os hiperenxebrismos, os lusismos, o cheísmo, o teísmo ou calquera outro fenómeno lingüístico terán cabida no corpus etiquetado como entidades propias que nalgún momento poderemos conectar coas variantes normativas a través dun hiperlema.

Nesta decisión pesan fundamentalmente dúas intencións. Por unha banda, o desexo de sermos fieis aos documentos orixinais. Pola outra, o rendemento do etiquetador e a semellanza entre os textos que conforman o corpus de adestramento e aqueles outros sobre os que se vai aplicar.

En primeiro lugar e polo que se refire ao primeiro dos aspectos, quen nos autoriza a manipular un documento escrito e publicado, con todos os dereitos que iso conleva, substituíndo nel palabras ou construcións utilizadas adrede polo seu autor polas correspondentes normativas, por moito que sexa con fins investigadores? En segundo lugar, como proceder cando sexan varias as posibles substitucións ou cando a magnitude do erro sexa tal que non se comprenda a intención do autor e non saibamos por que corrixir? En terceiro lugar, que beneficios nos achega maquillar e desfigurar a realidade lingüística galega contemporánea?

A resposta a estas tres preguntas é sinxela. O feito de non manipular de ningún xeito as secuencias dos diferentes documentos do corpus evítanos, ademais de calquera problema legal, o traballo de actuar de correctores, minimiza a problemática do etiquetado manual e permítenos evidenciar o estado real da lingua galega.

Se a todo anterior lle engadimos que o noso propósito é que o etiquetador se enfrente con toda a problemática derivada de tratar con textos reais, pois calquera outra cousa, aplicada ao galego, supoñería un apaño dificilmente defendible pola falta de practicidade e de realismo, non é de estrañar que sexamos respectuosos cos documentos que manexamos.

Agora ben, calquera das dúas opcións polas que poderíamos decidirmos seguirían sen contentar a todos os posibles usuarios do CORGA etiquetado, pois habería quen necesitase abstraerse de toda esa variedade gráfica reflectida tamén na lematización e

traballar só coas invariantes, mentres que outros usuarios precisarían poder acceder ás variantes. A solución pola que optamos inicialmente foi reflectir na etiquetaxe toda a heteroxeneidade ortográfica amosada nos documentos de traballo, para nun segundo momento poder tratala coa introdución no sistema dun hiperlema que agrupe as posibles variantes.

Pois ben, precisamente por non actuar de correctores, é fundamental que establezamos na etiquetaxe manual unha diferenza entre o tratamento das grallas e o de calquera outra desviación da norma. As primeiras son erros tipográficos, casuais, non recorrentes, mentres que os segundos son produto do descoñecemento lingüístico, de interferencias con outras linguas ou mesmo responden a unha escolla consciente. En xeral, as primeiras débense á supresión ou engadido dalgún carácter, provocando unha secuencia inexistente na nosa lingua; as segundas acostuman ser obxecto de estudos dialectolóxicos, sociolingüísticos e gramaticais.

4.3.1. O tratamento das erratas na etiquetaxe manual

A pesar de que nun primeiro momento mostrámonos partidarios de elevar as erratas outorgándolles estatus de palabras existentes, a semellanza de calquera outra do léxico galego, finalmente mudamos a nosa opinión debido ás implicacións que ese feito comporta.

A priori parece coherente e fácil de manter o criterio de respectar absolutamente a grafía de todas as unidades léxicas que aparecen nos textos a etiquetar. Mais logo, a medida que nos imos atopando cos distintos casos, comezan a asaltarnos as dúbidas. Ata que punto é correcto e aconsellable que etiquetemos unha forma e lle inventemos un lema baseándonos no erro tipográfico? Por exemplo, entre as erratas que aparecen na versión 2.4 do CORGAetq están as dúas seguintes:

Probablemente, os fluxos económicos dentro de Galicia, a acual rede de infraestructuras, a existencia da administración autonómica e dunha realidade territorial e administrativa galega, así como a propia conciencia colectiva, son os factores que explicarían a dirección desta apertura de oficinas e, en xeral, da estratexia das tres caixas. [CORGAetq: *GI1996-06/40*]

Xente que agracede que lle asomes a arte como un mero intercambio de reflexións e sentimentos sumidos na busca da fermosura. [CORGAetq: *GH2007-03-30/112*]

En ambos os casos, o lector percibe a primeira vista que nun *token* falta un *t* e que no outro se produce a permuta entre o *c* e o *d*, impedindo a aparición das unidades que deberían estar aí: *actual* e *agradece*.

Pois ben, sendo consecuentes co criterio de analizar e etiquetar o que figura no documento orixinal, realizamos as seguintes caracterizacións correspondentes ao adxectivo e á forma verbal:

actual:

```
<análise_unidade>
<unidade>actual</unidade>
  <constituínte>
    <forma>actual</forma>
    <etiqueta>A0fs</etiqueta>
    <lema>actual</lema>
  </constituínte>
</análise_unidade>
```

agradece:

```
<análise_unidade>
<unidade>agradece</unidade>
  <constituínte>
    <forma>agradece</forma>
    <etiqueta>Vpi30s</etiqueta>
    <lema>agraceder</lema>
  </constituínte>
</análise_unidade>
```

En principio todo parece correcto, mais o feito de etiquetar no corpus de adestramento do xeito anterior ten dúas consecuencias coas que non contabamos: unha, o etiquetador aprenderá que as etiquetas A0fs e Vpi30s se asocian respectivamente coas unidades *actual* e *agradece*, polo que, se volven concorrer, etiquetaraas correctamente; dúas, o etiquetador vai aprender que *actual* e mais *agraceder* son dous lemas máis do galego. En definitiva, tratará as dúas formas e os dous lemas coma se formasen parte do acervo léxico do galego e non os distinguirá de ningún xeito dos verdadeiros, xa que a frecuencia da unidade, malia que sexa de un, non vai servir para establecer diferenzas, posto que máis do 40% das unidades presentan unha frecuencia de un¹⁵⁰.

¹⁵⁰ A falta de datos concretos sobre o corpus etiquetado, sérvenos de mostra o CORGA: o 44,54% das 389 917 formas distintas que sen incluír as cifras se rexistran nese corpus presenta unha frecuencia de un.

En consecuencia, pensamos que debemos modificar esta circunstancia de xeito que poidamos seguir mantendo a fidelidade ao texto orixinal e, ao tempo, non terxiversar o léxico galego. A disposición e mais a estrutura do formato que escollemos para levar a cabo a desambiguación manual permítennos materializar este propósito.

Así, no interior das etiquetas xml *expresión* e *unidade*, que acollen respectivamente as distintas oracións e unidades léxicas presentes nun documento concreto, manteremos a unidade inalterada, é dicir, tal e como nos aparece na secuencia orixinal. Por exemplo, a voz *gastronómica*, documentada no GH2007-03-30/106¹⁵¹:

```

<parágrafo>
  <oración>
    <expresión>Para rematar a ruta gastronómica, unha paradiña no Restaurante París, onde nos
    levaremos á boca unha savorosísima pescada á cazola;</expresión>
    <análise>
      <análise_unidade>
        <unidade>Para</unidade>
          <constituínte>
            <forma>para</forma>
            <etiqueta>P</etiqueta>
            <lema>para</lema>
          </constituínte>
        </análise_unidade>
        <análise_unidade>
          <unidade>rematar</unidade>
            <constituínte>
              <forma>rematar</forma>
              <etiqueta>V0f000</etiqueta>
              <lema>rematar</lema>
            </constituínte>
          </análise_unidade>
          <análise_unidade>
            <unidade>a</unidade>
              <constituínte>
                <forma>a</forma>
                <etiqueta>Ddfs</etiqueta>
                <lema>o</lema>
              </constituínte>
            </análise_unidade>
          <análise_unidade>
            <unidade>ruta</unidade>
              <constituínte>

```

¹⁵¹ Non a subliñamos, como acostumamos destacar as formas que nos interesan, porque queremos que se vexa o texto tal cal aparece.

```

        <forma>ruta</forma>
        <etiqueta>Scfs</etiqueta>
        <lema>ruta</lema>
    </constituínte>
</análise_unidade>
<análise_unidade>
<unidade>gastronómica</unidade>

```

Mais, dado que o etiquetador prescinde totalmente da información relativa á *unidade* e, por suposto, tamén da de *expresión*, o feito de manter a errata nesas dous ámbitos non traerá consecuencias insospeitadas.

Non obstante, na análise da unidade léxica, ou sexa, nos tres parámetros que conforman a caracterización plena de calquera elemento gramatical –forma, etiqueta e lema– introduciremos os datos correctos:

```

<constituínte>
    <forma>gastronómica</forma>
    <etiqueta>A0fs</etiqueta>
    <lema>gastronómico</lema>
</constituínte>
</análise_unidade>

```

Do mesmo xeito, os valores correspondentes á forma e lema de *acual* e *agracede* que vimos máis arriba tamén serán corrixidos por:

```

acual:
    <análise_unidade>
    <unidade>acual</unidade>
        <constituínte>
            <forma>actual</forma>
            <etiqueta>A0fs</etiqueta>
            <lema>actual</lema>
        </constituínte>
    </análise_unidade>
agracede:
    <análise_unidade>
    <unidade>agracede</unidade>
        <constituínte>
            <forma>agradece</forma>
            <etiqueta>Vpi30s</etiqueta>
            <lema>agradecer</lema>
        </constituínte>
    </análise_unidade>

```

Na consideración de erratas entran non só formas inexistentes (*catoliscismo, catátrofe, claiificación, electrónico, enfervocidado...*) senón tamén aquelas outras que aínda existindo na nosa lingua é imposible que concorran en certos contextos, pois as súas características morfosintácticas non se adecúan ás da palabra que realmente debería aparecer. Un dos exemplos máis claros dáse na seguinte secuencia pertencente a un relato curto de *Remuíños en coiro*:

Caldo volveu, o Farruco, mostrando todo o seu orgullo de propietario, atopábase acororado enriba do seu libro máis preciado e pasaba as páxinas con semellante rapidez que non se vían as rachaduras que decoraban os folios de arriba abaixo, só se oían. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Nela, como observamos, a palabra inicial, o substantivo *caldo* foi introducido erroneamente polo adverbio relativo *cando*¹⁵².

Xa moito máis próximos no tempo á redacción deste apartado, atopámonos con que a tipoloxía de erros tipográficos se incrementa coa aparición de erros provocados pola acción non controlada de ferramentas automáticas como son correctores e tradutores. Novamente, trátase de unidades que forman parte do léxico galego e incluso son das máis utilizadas: *non, ou, fose, mocidade...* O problema é que non encaixan en absoluto no lugar no que concorren. Vexámolo:

Así mesmo, o portavoz socialista considera tamén "fundamental" a presenza de Salgado como ministra de Economía, clave para afrontar o "difícil contexto económico non que nos atopamos". [CORGA: *GH2009-04-09/79*]

Pérez Bouza replicou a Espiñenta que o Estado español pode acollerse a esta cláusula, destacou que "ou rexeitamento a este tipo de cultivos é case unánime, mesmo por parte de parlamentos como ou de Galiza e moitos concellos" e advertiu á ministra de que "fomentar ou permitir ou cultivo de transxénicos é claudicar perante unhas poucas multinacionais que ou único que buscan é controlar nun futuro próximo a totalidade do ciclo produtivo agrícola". [CORGA: *GH2009-04-30/73*]

Avisou ao comezo da segunda parte, con dous disparos, un de Ángel Lafita que despexou o meta brasileiro Diego Alves co peito e outro de Juan Rodríguez que se foi fose. [CORGA: *DLV2008-12-01/58*]

¹⁵² Para calquera cuestión relacionada coa clasificación categorial desta forma remitimos ao apartado da clase Adverbio do capítulo primeiro.

Vettel, que no G.P. de Italia do ano pasado arrebatou tres marcas de mocidade a Fernando Alonso _pole, vitoria e podio_, xogou moi forte na sesión de clasificación e gañou conseguir por segunda vez o primeiro posto da formación de saída. [CORGA: GH2009-04-19/87]

A explicación á aparición das dúas primeiras unidades subliñadas semella clara: nun primeiro momento escíbese ~ introdúcese ~ tradúcese¹⁵³ para *no* e *o*, e logo, nun segundo, aplícase con total seguridade un tradutor automático que transforma a suposta partícula negativa e a conxunción castelás nas correspondentes galegas *non* e mais *ou*. Naturalmente, o proceso final non foi controlado nin revisado por unha persoa, quen enseguida percibiría o erro.

Os dous casos seguintes comportan máis problemas. *Fose* debería ser *fóra* e *mocidade*, *novo*. O que realmente pasou dende o escrito inicial ao final só podemos intuilo en parte. O que é seguro é que *fuera* mudouse, ben manual ben automaticamente, primeiro por *fora* e deseguido por *fose*, ou xa directamente por este último, probablemente partindo da regra implícita de non permitir a aparición das formas en *-ra* como subxuntivas, fallando o tradutor ou equivocándose o escritor ao confundir o adverbio coa forma verbal, ou ao non grafalo ben. Ao último da serie destes rechamantes erros non conseguimos atoparlle explicación, pois aínda que a forma inicial fose *nuevo* ou improbablemente *joven*, chegaríamos a *novo* ou *mozo*, mais nunca ao abstracto *mocidade*; o que si é obvio é que “novamente” lle volveron arrebatarse as tres marcas a Alonso e que este feito non tivo lugar, precisamente, durante a súa “mocidade”.

A nosa proposta consiste en tratar os erros anteriores ou outros semellantes coma se fosen erros tipográficos e mantelos tal cal aparecen na expresión e na unidade, substituíndoos non obstante na forma e no lema, pois non ten sentido ningún incluír no lexicón, nin que o etiquetador adestre, a contracción *non*, o artigo *ou*, o adverbio *fose* ou a locución adverbial *de mocidade*. Así, a análise correspondente ás catro unidades anteriores sería:

non por *no*:

<análise>

¹⁵³ Non debemos ser inxenuos, e aínda que este feito non se explicita, na maior parte das noticias, sobre todo nas de axencia e moitas de redacción, pártese da versión castelá, aplícaselle un tradutor ao galego e logo revísase o resultado. É obvio en casos coma o da segunda secuencia, na que non só se ultracorríxese *o* por *ou*, senón na que tamén se traduce o apelido *Espinosa* por *Espiñenta*.

```
<análise_unidade>
<unidade>non</unidade>
<constituínte>
<forma>en</forma>
<etiqueta>P</etiqueta>
<lema>en</lema>
</constituínte>
<constituínte>
<forma>o</forma>
<etiqueta>Ddms</etiqueta>
<lema>o</lema>
</constituínte>
</análise_unidade>
```

ou por o:

```
<análise>
<análise_unidade>
<unidade>ou</unidade>
<constituínte>
<forma>o</forma>
<etiqueta>Ddms</etiqueta>
<lema>o</lema>
</constituínte>
</análise_unidade>
```

fose por fóra:

```
<análise>
<análise_unidade>
<unidade>fose</unidade>
<constituínte>
<forma>fóra</forma>
<etiqueta>Wn</etiqueta>
<lema>fóra</lema>
</constituínte>
</análise_unidade>
```

de mocidade por de novo:

```
<análise>
<análise_unidade>
<unidade>de mocidade</unidade>
<constituínte>
<forma>de novo</forma>
<etiqueta>La0</etiqueta>
<lema>de novo</lema>
</constituínte>
</análise_unidade>
```


Deste xeito solventamos na etiquetaxe manual os problemas inherentes aos erros tipográficos e, en consecuencia, o modelo de adestramento lingüístico que propoñemos é coherente e consistente.

Na etiquetaxe automática, non obstante, imos atoparnos con resultados diverxentes.

O etiquetador de ningún xeito é quen de detectar nin de corrixir unha errata –e tampouco é ese o seu cometido–, polo que ante os erros tipográficos vaise comportar como se non o fosen e vai analizar o texto que se lle proporciona. Posto que un acento mal colocado ou unha letra que sobra ou falta converten a palabra pretendida noutra, existente ou non na nosa lingua, o etiquetador vai actuar de modo distinto segundo o caso. Se a unidade resultante coincide con outra que teña no lexicón, outórgalle a etiqueta e o lema que a el lle consta; se non a atopa, non lle vai asignar ningún lema.

É dicir, cando o erro tipográfico dea lugar a unha unidade para a cal o etiquetador non ten ningunha información (por exemplo, *esoite* en vez de *escoite*), trátaraa coma se estivese ante unha palabra descoñecida, pois para el éo, e asignaralle unha etiqueta en función da similitude da terminación da errata coa doutras formas que ten no lexicón, das formas e etiquetas que están antes e despois, da estatística, etc., pero non lle asignará lema, no seu lugar aparecerá un asterisco¹⁵⁴.

Cando o erro tipográfico provoca a coincidencia cunha forma xa existente no lexicón, por exemplo a que viamos máis arriba de *caldo* en vez de *cando*, o etiquetador automático outorgará a que a el lle consta: Scms (*substantivo, común, masculino, singular*).

4.3.2. O tratamento das desviacións da norma

Ante a problemática de establecer unha liña divisoria do que se considera errata fronte a calquera das desviacións ou vacilacións que veremos de seguido, preferimos pecar por defecto que non por exceso. Deste xeito, calquera irregularidade ou incorrección dende o punto de vista da norma susceptible de atribuírse a un emprego consciente por parte do autor, á diversidade dialectal do galego ou a cuestións

¹⁵⁴ Marca coa que escollemos sinalar na etiquetaxe automática a ausencia do lema no lexicón. Podería ter sido calquera outro signo o elixido, mais o importante é marcar dalgún xeito esa ausencia, de modo que nos resulte cómodo poder recuperar a información relativa ás unidades que non están contidas no lexicón.

relacionadas coas escollas normativas como son til diacrítico, xénero dalgúns substantivos, acentuación conflitiva, hiperenxebrismos, etc. manteranse na casilla *forma* e proporcionaráselles o lema que corresponda a cada unha desas ocorrencias.

Dada a imposibilidade de afrontar polo momento a adición no sistema XIADA dun campo hiperlema, é dicir, aquel que agrupe ao seu redor as múltiples variantes gráficas dunha unidade, optamos por que se reflecta na etiquetaxe, tanto manual como automática, toda a variedade que amosa o galego, de xeito que cada unidade ortográfica na que, exceptuando os morfemas flexivos, se produza algunha variación con respecto á forma canónica constituirá para o sistema, e para nós, unha unidade de análise distinta. Vexamos primeiro as seguintes secuencias e logo a etiquetaxe do adxectivo subliñado e visualizarase facilmente o que pretendemos explicar.

Cecais a débeda histórica de comunidades tradicionalmente deficitarias, como a galega, poida calcularse así como a diferenza entre o custe efectivo e os recursos necesarios para soste un nivel de servizos aceitábel, en áreas como a educativa ou a sanitaria, desde as transferencias das mesmas ata hoxe. [CORGAetq: *GI1996-05/55*]

Suso de Toro sinala que no apartado da tradución estamos lonxe de alcanzar un nivel acetable. [CORGA: *VG1997-11-15/8*]

Creo que flexibiliza-lo mercado laboral é o camiño acceptable para que a nosa economía adquira un nivel de competitividade parecido ó dos estados do noso contorno. [CORGAetq: *CG1999-05-09/9*]

A Fundación March encargouse da súa restauración e fixación, hoxe atópase a maioría delas nun estado de abondo aceptábel. [CORGAetq: *ANT2005-03-03/74*]

Como vemos, recollemos unha ocorrencia de catro maneiras diferentes de grafar o mesmo adxectivo. Pois ben, malia que a segunda non foi aínda etiquetada, o que teríamos, coa etiquetaxe que practicamos e defendemos, son catro unidades distintas, cada unha delas cunha forma e un lema:

aceitábel	A0ms	aceitábel
acetable	A0ms	acetable
acceptable	A0ms	acceptable
aceptábel	A0ms	aceptábel

Este mesmo principio aplícase en todas as categorías e a todas as unidades, incluso nos nomes propios. Por exemplo, as tres maneiras en que aparece escrito o

nome dun grupo musical –non deixa de ser curioso que esta variación teña lugar na mesma noticia–, ou o intento de adaptación mediante o til de topónimos estranxeiros, teñen como consecuencia a existencia de tantas unidades de análise diverxentes como formas de escribilas se documentan.

Así, as formas subliñadas das seguintes cinco secuencias

Mutenrohi, Xochilmica, Bonovo, Xarín e Uxía Senlle estrean cadanseu novo disco nos palcos e nos festivais que ateigan o calendario galego [CORGAetq: ANT2008-07-10/48]

Tamén sobre o escenario se medirá a acollida do xiro que vén de dar Xochimica, no seu segundo e agardado disco, Carta Blanca. [CORGAetq: ANT2008-07-10/48]

Nova pel de Xochimilca. [CORGAetq: ANT2008-07-10/48]

Non é que a democracia ou os dereitos humanos brillen polo seu esplendor na Rusia de Putin, pero o desentendemento entre as dúas administracións obedecerá, cabe pensar, a outras causas, como outras causas, e non as oficiais, motivan a recente escalada contra Siria, a quen, por certo, Moscova anunciou que vai vender sistemas antiaéreos Strelets, contra o parecer de Washington, que, tamén, está en contra do levantamento do embargo de armas a Beijing, porque o que pretende é venderlle armamento en exclusiva a todo o mundo. [CORGAetq: ANT2005-03-03/65]

Dan Miller, portavoz da empresa Majestic America, de Seattle, en Wáshington, dixo que o buque navegaba no segundo día dun cruceiro dunha semana cando, por razóns que non se aclararon, chocou contra unha roca no extremo sur do estreito Icy, onde as augas teñen unha profundidade dun metro. [CORGAetq: DLV2007-05-15/68]

dan lugar a outras tantas formas e lemas distintos:

Xochilmica	Sp00	Xochilmica
Xochimica	Sp00	Xochimica
Xochimilca	Sp00	Xochimilca
Washington	Sp00	Washington
Wáshington	Sp00	Wáshington

Polo que se observa, compróbase que a efectos de etiquetaxe non nos importa a normatividade ou non normatividade da forma nin do lema e tamén que non establecemos diferenzas en función desta; o noso obxectivo final é facilitar a extracción de información e as consultas sobre un corpus etiquetado no que se recolla e trate toda a variedade lingüística existente nos documentos orixinais. Naturalmente, o noso

desiderátum é poder traballar tamén con hiperlemas, de xeito que poidamos localizar todas as variantes dunha unidade mediante unha única consulta, mais polo de pronto non é factible.

Vistos os exemplos de variación anteriores, procederemos agora a amosar a solución adoptada na etiquetaxe manual dos varios tipos de vacilacións ou erros ortográficos/lingüísticos que, vistos sempre dende a norma, provocan problemas de distinta índole no etiquetador.

4.3.2.1. Vacilacións ~ variacións no vocalismo

Polo que respecta ás variacións no vocalismo imos destacar dous tipos, segundo a alternancia se dea na clase verbal ou en calquera outra, en función do elevado número de elementos que posúe o paradigma da primeira fronte ás outras e que, polo tanto, repercute nunha cantidade maior de formas:

1) Categoría non verbal. Xeralmente, alterna a forma con vogal de grao medio e a forma con vogal pechada. Exs.:

Pola inmensa sorte do éxito do noso sistema social, hoxe, progresivamente e desde mediados os sesenta en que empezamos a meter o fociño a clase media, millóns de persoas gozan do sol e da area das praias. [CORGAetq: GH2008-08-05/98]

Fontes do Grupo de Rescate e Estudo de Mamíferos Mariños informou de que o golfinho mular (*Tursiops Truncatus*) ten unha ferida no fuciño, un corte vertical desde o interior da boca cara á fazula, e, aínda que se trata dunha "zona delicada", onde recibe sons, non reviste gravidade. [CORGAetq: GH2008-08-05/41]

Outras comparecencias que tamén serán críticas coa xestión da catástrofe, e que terán lugar ese mesmo día serán as de Fiz Fernández, do Instituto de Investigacións Mariñas, e Martiño Nercellas, da Federación Ecoloxista Galega. [CORGAetq: ANT2005-03-03/23]

Por exemplo, o venres da semana pasada, un informe do Banco de Investimento americano Goldman Sachs afirmando que o sector bancario europeo ía necesitar 90.000 millóns de euros para capitalizar os seus balances, provocou nese mismo día unha perda no valor das accións dos principais bancos europeos dun 3% de media, sen discernir practicamente bancos españois de bancos ingleses ou bancos grandes de bancos pequenos. [CORGAetq: ANT2008-07-10/72]

Nestes dous casos concretos, o etiquetador non ten problemas para recoñecer as unidades, pois ambas constan no lexicón por estar entre as 100 000 formas máis frecuentes do CORGA. O seu tratamento na etiquetaxe manual responde ao principio

establecido de respectar a caracterización e lematización das variantes. Así, figuran na listaxe de formas, etiquetas e lemas os dous pares como entidades distintas:

fociño	Scms	fociño
fuciño	Scms	fuciño
mismo	Idms	mismo
mesmo	Idms	mesmo

2) Categoría verbal. Neste subtipo debemos distinguir entre os verbos nos que se produce unha variación similar á comentada no tipo 1 daqueles outros que posúen na súa conxugación alternancia vocálica na raíz, e finalmente os que cambian a súa adscrición a unha das tres clases fronte á conxugación que determina a normativa.

Polo que se refire ás variacións no vocalismo semellantes ás sinaladas no apartado 1, acostuman ser verbos da primeira conxugación regular, aínda que tamén se dá algún caso de variación noutras clases, polo que o etiquetador automático adoita atinar coa etiqueta, malia que non proporcione o lema por non figurar no lexicón. Na etiquetaxe manual conforman unha análise diferente da do verbo normativo, estando en consonancia o lema coa forma gráfica que concorre. Por exemplo,

"Non se lle cocía o pan no corpo a Don Quixote, como adoita dicirse, ata ouvire e saber as maravillas prometidas do home portador de armas", arrinca a tradución do capítulo 25, "onde se aponta a aventura do orneo e a chusqueira do titereiro, coas lembradas adiviñanzas do macaco adiviñador". [CORGAetq: ANT2005-03-03/33]

aponta	Vpi30s	apontar
--------	--------	---------

fronte a:

O peor é que nin para o 2008 estarán listas", apunta Rodríguez. [CORGAetq: ANT2002-10-03/31]

apunta	Vpi30s	apuntar
--------	--------	---------

Outros exemplos do mesmo tipo son os seguintes:

Nun mes, tócalle a vez a Rajoy, supoño que alguén lle vai preguntar polos 80 céntimos do café aínda que no seu caso, queda mellor falar do que costa un puro... [CORGAetq: GH2007-03-30/100]

O maior inconveniente das lámpadas de baixo consumo é o seu elevado prezo, xa que agora custan até sete veces máis que as antigas. [CORGAetq: ANT2008-07-10/55]

A día de hoxe, a sensibilización adquerida sobre este patrimonio tamén chega ao ámbito político e así desde o concello de Mondariz e de Covelo téñense erguido voces (BNG) que demandan a catalogación e protección dos foxos de Gargamala e Sabaxáns (documentados en 1664) situados no Chan da Anduriña e ameazados, xunto cunhas mámoas, pola abertura de dúas canteiras e un parque eólico. [CORGAetq: ANT2008-07-10/59]

Os países africanos reclaman o cumprimento destas axudas, calculadas en 50.000 millóns de dólares para o 2010, e que están enmarcadas dentro do compromiso adquirido polas principais potencias de asumir os 'Obxectivos do Milenio da ONU', estipulados para o 2015. [CORGAetq: ANT2008-07-10/79]

No tocante á organización do poder galego, a proposta presenta varias novidades no que se refire sobre todo ao poder xudicial, concebido arredor dun Consello Xudicial de Galiza que represente o órgano de goberno dese poder en Galiza. [CORGAetq: ANT2005-03-03/2]

Desexando manter vivo o espírito inicial do proxecto (pois a empresa "fora concibida por Ibáñez como unha asociación de recursos propios para, engadíndolles valor, poder satisfacer necesidades auténticas, o que viña significar unha clara revolución fronte á empresa que xeralmente está concibida como unha asociación de intereses co fin de especular con necesidades e recursos"), nace _ó abeiro do Laboratorio de Formas de Galicia_ un complexo, dinámico e ambicioso plan de restauración de Sargadelos. [CORGAetq: CG2000-12-03/77]

estes implican a coexistencia no mesmo sistema das análises seguintes, sen que existan ligazóns de ningún tipo entre os distintos membros do par nin marcas que releguen un e prioricen outro:

costa	Vpi30s	costar
custan	Vpi30p	custar
adquerida	V0p0fs	adquerir
adquirido	V0p0ms	adquirir
concebido	V0p0ms	concebir
concibida	V0p0fs	concibir

Nos casos de conxugación verbal en que a característica xeral é a alternancia vocálica no radical no presente de indicativo, como sucede con *servir* ou *durmir*, cando non se realiza dita alternancia o etiquetador automático erra a análise, pois asígnalle a que corresponde á forma gráfica que lle consta no lexicón. Por exemplo, ante

O conflito evidente entre civilizacións sirve de base, en parte, para xustificar unhas actuacións nas que está en xogo o liderazgo global no medio prazo. [CORGAetq: ANT2002-09-12/31]

o etiquetador, por non figurar no lexicón a forma *sirve* caracterizada como terceira persoa do presente de indicativo, asigna aquela que coñece, a de segunda persoa do singular do imperativo. Naturalmente, neste caso concreto nós corriximos na etiquetaxe manual esta última etiqueta pola primeira, de xeito que no corpus de adestramento, igual ca en parte do territorio galego, conviven para a terceira do singular do presente de indicativo do verbo *servir* dúas formas, unha con alternancia –a normativa e máis frecuente no CORGAetq, con 10 ocorrencias– e a outra sen alternancia –a dialectal e máis inusual no corpus, cunha soa ocorrencia–:

sirve	Vpi30s	servir
serve	Vpi30s	servir

O mesmo tratamento pola nosa parte reciben os *tokens resintén* e *sufren* presentes en

os que aforran carga impositiva son os máis ricos, os servicios públicos resíntense, as desigualdades sociais aumentan e a dinámica da conxuntura económica non se ve afectada por estas medidas nun sentido predecíbel e favorábel. [CORGAetq: ANT2002-11-07/29]

Pola vía dos gastos as operacións de capital sufren un enorme recorte cunha caída do 13'2 por cento, en tanto que aumentan máis que proporcionalmente as operacións correntes, sobre todo, os gastos en bens e servicios correntes e os gastos financeiros. [CORGAetq: GII996-01/32]

resintén	Vpi30p	resentir
sufren	Vpi30p	sufrir

coa diferenza con respecto a *sirve* de que o etiquetador automático xa proporciona as etiquetas correctas, a de terceira do plural do presente de indicativo, posto que non hai ningún elemento que lle avolva esas análises por non darse sincretismo formal con ningunha outra forma verbal do paradigma de *consentir* e *sufrir*. É dicir, fronte a *sirve* que na conxugación sen alternancia desempeña dous valores –o de presente de indicativo e o de imperativo–, as formas da segunda singular e da terceira plural do presente indicativo, malia conxugarse sen alternancia, non coinciden formalmente con outras do seu mesmo paradigma, razón pola que o etiquetador, seguindo as pautas das palabras descoñecidas, as etiqueta ben aínda que non lles outorga lema.

Finalmente, non é inusual atopar nos textos ocorrencias de formas verbais que se clasifican nunha conxugación distinta da que impón a normativa, adaptacións de verbos

casteláns ou portugueses ou incluso emprego destes por influencia desas dúas linguas. Unha mostra destes usos vémosla nos seguintes exemplos nos que recolleemos en xeral un caso de cada par –o normativo e unha variante non normativa–:

Os dous candidatos están a favor de reducir o uso de combustibles fósiles para combater o cambio climático. [CORGAetq: GH2008-08-05/66]

Tamén é un medio bo para combatilo, fomentar as nosas pequenas ambicións, loitar para realizalas cabalmente e amar o esforzo que elo require. [núcleo de XIADA]

Barreiro dixo que non sería difícil conseguir algún crédito para estas obras e o propio José Luís Méndez dixo que si. [CORGAetq: GH2007-03-17/44]

Segundo os seus cálculos, Brasil terá que conseguir 63 mil millóns de dólares en 2003 para evitar un cesamento de pagos. [CORGAetq: ANT2002-10-31/27]

Coa reunión de ministros do próximo día vintecatiro, a sombra da reestructuración volve cernerse sobre Astano e sobre Galicia. [CORGAetq: TN1997-04/7]

O baruto, tiña, a trama moi fina, para cernir a "pintura" ou lista, e tamén o "alcohol de folla" cando aínda non viña preparado e precisábase pisalo e cribalo. [CORGA: A. Abella e H. M. Pose, *Alfarería de torno en Buño*, Escola-Taller Casa do Oleiro, 1994]

Incluso hai prognósticos feitos hai moito tempo sobre temas demográficos, do campesiñado, da economía mariñeira, prognósticos dos riscos que se cernían sobre estas realidades que se están cumprindo como se non existise un poder político propio, de autogoberno e a min paréceme que iso é gravísimo. [CORGAetq: GH2008-08-05/43]

En definitiva, cínguense máis a realidades e viven máis o curto prazo. [CORGAetq: GH2008-08-05/31]

Só se cinxe a defende-los intereses de Erimsa, se ben reconece que "existe un bo contrato" cos noruegueses. [CORGAetq: CG1996-02-07/38]

Unha delas é que cando se rescata a concesión antes dese prazo, hai que indemnizar e iso permite obter recursos para un novo emprazamento. [CORGAetq: ANT2005-03-03/19]

O pacote de rescate financeiro de 30 mil millóns de dólares para Brasil anunciado en agosto polo FMI, está en realidade destinado a resgatar os principais acreedores de Brasil, como o Citibank, J.P. Morgan e outros grandes bancos internacionais. [CORGAetq: ANT2002-09-12/34]

A acusación baseábase na discriminación sufrida por "razón de sexo", xa que, na súa opinión, o seu salario "reducírase ó non beneficiarse dos aumentos acordados con efectos retroactivos". [CORGAetq: CG1996-02-14/119]

Ousaría dicir que as demandas catalanas, que se basan nunha concepción política do que debe ser o estado autonómico, teñen a virtude de poñer ás autonomías diante das súas propias responsabilidades. [CORGAetq: *GII1996-06/41*]

A variación nun mesmo verbo pode aínda incrementarse máis se aos cambios de terminación, e polo tanto de clase conxugacional, se lle une a problemática da variación na vogal radical. Dáse isto, por exemplo, no principal verbo de dicción galego, *dicir*, para o que amais da variante dialectal correspondente á segunda conxugación, *decer*, aparece tamén nos textos o castelanismo *decir* e o arcaísmo ou probable lusismo *dicer*:

Quere iso dicir que os xornais están cheos de "novelas contemporáneas"? [CORGAetq: *ANT2008-07-10/27*]

Un dos problemas das fábricas ó sulfato é o da polución á cor, é decer, os verquidos chegan a coller unha cor, un tinte moi escuro contra do que actualmente non existen tratamentos eficaces. [CORGAetq: *TE1977-10/4*]

Non me atrevía a lle decer á miña nai que falara con aquel home. [CORGA: X. L. Méndez Ferrín, *Arraianos*, Xerais, 1991]

Un corvo chamoulles a atención facendo circos no ceu, como querendo dicer algo na lingua dos tempos. [CORGA: L. Villalta, *As chaves do tempo*, A Nosa Terra, 2001]

Isto provoca a aparición na etiquetaxe de catro lemas distintos para a expresión da mesma realidade:

dicir	V0f000	dicir
decir	V0f000	decir
decer	V0f000	decer
dicer	V0f000	dicer

Algo semellante ocorre tamén con outro verbo da segunda conxugación, *caer*, para o cal, á parte da variante dialectal que sofre o cambio de conxugación, *caír*, documéntase aínda outra máis, mutación desta última por diferenzas na acentuación, *cair*. Un exemplo de cada un deles vémolos en

Probabelmente todo comezou cando Putin botou man de Lukos para impedir que esta empresa estratéxica viñese caer nas redes estadounidenses. [CORGAetq: *ANT2005-03-03/65*]

non podemos permitir a desgaleguización de Galiza, o esforzo que nos levou a plebiscitar un Estatuto en 1936 non pode caír en saco roto, busquemos a maneira de erguer unha bandeira que nos represente. [CORGAetq: *ANT2008-07-10/4*]

A como están as cousas, Brasil podería caír no cesamento de pagos aínda antes de que tome posesión o novo presidente en xaneiro, cando, como en Arxentina en 2001, non haberá reservas para pagar a débeda. [CORGAetq: ANT2002-10-31/27]

Esta variedade nos textos obríganos a considerar tres lemas distintos, sen posibilidade polo momento de abstraelos a un hiperlema:

caer	V0f000	caer
caír	V0f000	caír
cair	V0f000	cair

Por suposto, a vacilación na acentuación non é casual e tampouco podemos achacala a un erro tipográfico, tal como nos demostran as estatísticas dos dous infinitivos no CORGA¹⁵⁵, con abonda presenza nos textos dende o ano 1975 ata o 2004:

CAÍR

Estatísticas por medio		Estatísticas por área temática		Estatísticas por lustro	
Xornais	<u>6</u>	Economía e Política	<u>5</u>	1975-1979	<u>12</u>
Revistas	<u>3</u>	Cultura e Artes	<u>1</u>	1980-1984	<u>24</u>
Libros	<u>94</u>	Ciencias Sociais	<u>16</u>	1985-1989	<u>6</u>
		Ciencias e Tecnoloxía	<u>2</u>	1990-1994	<u>22</u>
		Ficción	<u>83</u>	1995-1999	<u>19</u>
		Outros	<u>3</u>	2000-2004	<u>19</u>
				2005-2009	<u>1</u>
				Total: 103	

CAIR

Estatísticas por medio		Estatísticas por área temática		Estatísticas por lustro	
Xornais	<u>8</u>	Economía e Política	<u>21</u>	1975-1979	<u>79</u>
Revistas	<u>3</u>	Cultura e Artes	<u>2</u>	1980-1984	<u>80</u>
Libros	<u>210</u>	Ciencias Sociais	<u>28</u>	1985-1989	<u>41</u>
		Ciencias e Tecnoloxía	<u>5</u>	1990-1994	<u>4</u>
		Ficción	<u>181</u>	1995-1999	<u>3</u>
		Outros	<u>14</u>	2000-2004	<u>14</u>
				2005-2009	<u>0</u>
				Total: 221	

¹⁵⁵ Dado que no CORGA non son posibles as buscas por lema, restrinximos a consulta á forma do infinitivo, polo que é probable que, de ter en conta todo o paradigma, o número de casos dunha e outra variante se incrementen.

Continuando con variacións no vocalismo frecuentes nos textos, é tamén digna de mención a transformación de hiato a ditongo, tanto na vogal da serie palatal como na da serie velar en posición final cando son átonas. Así:

oe > oi

doe > doi

cae > cai

sae > sai

ceo > ceu

grao > grau

Non debemos pensar que este fenómeno é só usual nos textos previos á primeira normativa oficial do galego, pois case todos os casos listados hai un momento teñen presenza en documentos posteriores ao 1982, chegando incluso a concorrer en pleno século XXI.

Vexamos algúns exemplos deste fenómeno típico da fala trasladado á escrita:

Sentíame bastante mal, cona, oise el mesmo dicir, e estírase con estas palabras axeitadamente na ancha cama de matrimonio. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Para quen coñeceu unha vez a conmoción dunha muller sabe que lembrala doi, mais sabe tamén que ese é o único modo de a resucitar, de resucitar nel a paixón que xamais desaparece do seu peito. [CORGA: R. Raña, *O crime da rúa da moeda vella*, Xerais, 1993]

Hoxe en día non se comen máis que asadas ou cocidas, pero os nosos devanceiros facían caldo de castañas ou de maias; acompañábanas con leite; botábanllas ó caldo de nabizas ou de berzas; e tamén as comían cocidas logo de salgalas coa tona na salmoira que cai do baño da carne do porco. [CORGA: *CG2000-11-11/41*]

Quedou o ceu vacío de nubes e paxaros, un ceu vacío de esperanzas. [CORGA: X. M. Martínez Oca, *Como paxaro caído do niño*, Baía, 2006]

Os moralistas admíranse do grau de venalidade, e o pobo humilde de que se poda pendurar un Miró na sala de baño (o pobo miúdo non lle daría corte a Miró se non viñese mesturado con personaxes de *Hola e Semana*). [CORGA: *TN2006-05/37*]

As análises que lles corresponden a esas variantes son, en consonancia coa nosa proposta, as seguintes:

oi Vpi30s oír

doi	Vpi30s	doer
cai	Vpi30s	caer
ceu	Scms	ceu
grau	Scms	grau

que conviven no corpus coas súas homólogas normativas:

oe	Vpi30s	oír
doe	Vpi30s	doer
cae	Vpi30s	caer
ceo	Scms	ceo
grao	Scms	grao

4.3.2.2. Vacilacións ~ variacións no consonantismo

4.3.2.2.1. Flutuacións B / V

O maior respecto cara á etimoloxía amosado no galego en voces como *avogado* (<ADVOCATUM), *voda* (<VOTAM), *vogar* (<VOCARE), etc. contrasta coa innovación do castelán (*abogado*, *boda*, *bogar*...) que, en última instancia, provoca que os escritores en lingua galega, na súa maioría alfabetizados e escolarizados en castelán, non sempre sexan conscientes desta diverxencia entre as dúas linguas. Esta interferencia é a única explicación plausible para a aparición nos textos, á parte das voces anteriores, de *maravilla* e familia (*maravillar*, *maravilloso*, *maravillosamente*), *móvil* e derivados (*automóvil*, *movilidade*, *movilizar*, *movilización*), *bulto*, *cobarde*, *esbelto*, etc., etc.

En consonancia co criterio establecido de respecto cara ao texto orixinal e o desexo de que a etiquetaxe amose toda a diversidade ortográfica que caracterizou a nosa lingua (e da que por desgraza aínda atopamos máis mostras das que nos gustaría) na etiquetaxe manual que desenvolvemos dámoslles cabida a ambos os pares. Así, no corpus de adestramento figuran por exemplo

Notemos que o proceso de redución das desigualdades interrexionais en España estancouse desde inicios da década dos 80, coincidindo co impacto da crise económica _que reduce a movilidade espacial da poboación_ e a maior liberalización exterior _que previsibelmente actúa como elemento de agudización dos desequilibrios_. [CORGAetq: *GI1996-05/55*]

Convén lembrar que este tipo de imposicións abusivas dana especialmente as posibilidades laborais ou sociais dos máis desfavorecidos, recortando as súas alternativas e a súa movilidade. [CORGAetq: ANT2008-07-10/17]

"Non se lle cocía o pan no corpo a Don Quixote, como adoita dicirse, ata ouvire e saber as maravillas prometidas do home portador de armas", arrinca a tradución do capítulo 25, "onde se apunta a aventura do orneo e a chusqueira do titereiro, coas lembradas adiviñanzas do macaco adiviñador". [CORGAetq: ANT2005-0303/33]

Os produtores fan unha chamada a todas as persoas que teñan nos seus arquivos persoais unha grabación, unha foto, un debuxo ou calquera outro documento visual ou sonoro de Manuel María. [CORGAetq: ANT2005-0303/37]

Hoxe coas técnicas que hai na gravación fánse marabillas, pois se alguén grava algo no norte do Canadá todos o coñecemos de seguido. [CORGAetq: ANT2005-03-03/49]

que se etiquetan respectivamente como

movilidade	Scfs	movilidade
movilidade	Scfs	movilidade
maravillas	Scfp	maravilla
grabación	Scfs	grabación
gravación	Scfs	gravación
marabillas	Scfp	marabilla
grava	Vpi30s	gravar

A diferenza doutras unidades incorrectas en galego que non se recollen no lexicón, por exemplo os verbos en *-uir*, as caracterizadas pola oscilación *b/v* si teñen presenza nel, pois, á súa vez, aparecen recollidas, a maioría, no VOLGa, malia que como formas a rexeitar.

Na etiquetaxe automática, cando a forma a analizar non figura nin no lexicón nin no corpus de adestramento, unha vez máis, a terminación e os pares forma-etiqueta anteriores e posteriores xogan unha baza determinante no recoñecemento. Un exemplo de aplicación do etiquetador automático sobre dúas formas descoñecidas para el son¹⁵⁶:

¹⁵⁶ O descoñecemento demóstrase na lematización coa aparición do asterisco.

Oración:	A mobilización contou con miles de persoas.							
Formas:	a	movilización	contou	con	miles	de	persoas	.
Etiquetas:	Ddfs	Scfs	Vei30s	P	Scmp	P	Scfp	Q.
Lemas:	o	*	contar	con	mil	de	persoa	.

Oración:	Teño un móvil novo!				
Formas:	teño	un	móvil	novo	!
Etiquetas:	Vpi10s	Dims	Scms	A0ms	Q!
Lemas:	ter	un	*	novo	!

4.3.2.2.2. Adverbios en -mentes

É bastante frecuente nos textos a aparición de adverbios rematados en *-mentes*, supoñemos que por un prurito de diferencialismo, procurando fuxir así da coincidencia co castelán. Na nosa opinión constitúen un exemplo máis do descoñecemento da normativa do galego actual ou ben unha escolla consciente por parte do autor do texto.

Un exemplo deste uso e da etiquetaxe que con eles practicamos podemos velo en

Ha de recoñecerse certa habilidade aos que manexan a axenda en inducir no cidadán temores e problemas que dificilmente lle afectarán directamentes e que, en calquera caso, son moi difíciles de solucionar de maneira estable e duradeira con políticas agresivas. [CORGAetq: ANT2002-09-12/31]

```

<análise_unidade>
  <unidade>directamentes</unidade>
    <constituínte>
      <forma>directamentes</forma>
      <etiqueta>Wn</etiqueta>
      <lema>directamentes</lema>
    </constituínte>
  </análise_unidade>

```

Xa Carballo Calero dá conta na súa *Gramática elemental* de que

La forma *-mentes* del sufijo fue preferida por algunos escritores en los años veintes: *somentes* ‘solamente’. Era el prurito del diferencialismo, como mecanismo de defensa contra la castellanización; pero tenía su fundamento en la realidad del habla, que no sólo ofrece

ejemplos de s paragógica en los adverbios terminados en *-mente*, sino en otras formaciones.
[Carballo, 1976: 240]

Freixeiro (2000: 517) límitase a subliñar a expresividade do sufixo *-mente* xunto co *-s* analóxico, mentres que Hermida (2004: 153) só comenta a incorrección de formas como *somente* e *somentes* e, finalmente, non o mencionan Álvarez, Regueira e Monteagudo (1986: 443-444) nin Álvarez e Xove (2002: 599-600).

Resúltanos sumamente curioso, e téntanos a profundar no seu estudo no futuro, o feito de que de todas as ocorrencias de adverbios en *-mentes* obtidas da versión 1.5 do CORGA, a ampla maioría aparece só unha vez, cinco concorren en dúas e unicamente un deles supera as cincocentas ocorrencias. Cremos que pode ser de interese ver cales son os adverbios que se documentan e cal é a súa frecuencia absoluta, así como a posición que cada un deles ocupa na listaxe completa de formas únicas do CORGA, motivo polo que a continuación amosamos esa información:

Posición	Forma	Frecuencia
4472	somentes	522
81175	soamentes	9
117417	sómentes	5
176806	simplesmentes	2
183933	casementes	2
185990	precisamentes	2
200856	novamentes	2
207029	talmentes	2
256123	todamentes	1
265022	exaseradamentes	1
267919	recentementes	1
276327	razoablementes	1
280367	directamentes	1

Posición	Forma	Frecuencia
283920	tristemementes	1
285603	certamentes	1
286472	irremisibelmentes	1
289586	expresamentes	1
294495	verdadeiramentes	1
294787	particularmentes	1
310181	fundamentalmentes	1
322466	absolutamentes	1
345008	escasamentes	1
352420	pesadamentes	1
363765	desesperadamentes	1
374371	desgraciadamentes	1
381270	lóxicamentes	1

Curiosamente, o adverbio en *-mente* máis frecuente da versión 1.5 do CORGA non é *soamente*, senón que este ocupa a quinta posición precedido por:

Posición	Forma	Frecuencia
687	especialmente	3302
692	precisamente	3290
888	realmente	2652
916	finalmente	2578
919	soamente	2569

Agora ben, se contabilizáramos as ocorrencias das seis variantes que concorren para este adverbio e puideramos subsumilas baixo un hiperlema, atopariámonos con que, efectivamente, é o máis empregado, chegando a sumar 3410 aparicións:

919	soamente	2569
4472	somentes	522
7266	somente	303
81175	soamentes	9
117417	sómentes	5
185597	sómente	2

No recoñecemento automático a presenza da secuencia *-mente*, con ou sen *-s* paragóxico, garante un correcto recoñecemento do adverbio, malia que este non estea rexistrado no lexicón nin figure no corpus de adestramento, como é o caso de *esveltamentes*:

Oración:	A estatua amosa unha boca e pescozo esveltamentes traballados.									
Formas:	a	estatua	amosa	unha	boca	e	pescozo	esveltamentes	traballados	.
Etiquetas:	Ddfs	Scfs	Vpi30s	Difs	Scfs	Cc	Scms	Wm	A0mp	Q.
Lemas:	o	estatua	amosar	un	boca	e	pescozo	*	traballado	.

4.3.2.2.3. Simplificación de grupos consonánticos cultos

Non é a nosa intención tratar neste apartado a simplificación dos grupos *-ct/-cc-* regulamentada na última edición das *Normas* (RAG/ILG, 2003), senón simplemente mostrar como na escritura, en concreto nos textos do CORGA, se reflicte a tendencia predominante na oralidade de redución dos grupos consonánticos. A solución adoptada na etiquetaxe que practicamos vai coincidir coa mencionada antes, pois caracterizaranse

as formas con grupo reducido ou complexo segundo aparezan, e lematizarase tamén en consonancia coa expresión da unidade gráfica.

A simplificación pode producirse pola eliminación da primeira das consoantes do grupo –debido á tendencia natural na oralidade da que falabamos antes, polo influxo do castelán que non posúe eses grupos nas mesmas palabras ou por descoñecemento–:

Todas as rutas de trasporte escolar terán acompañante [CORGAetq: GH2007-03-30/28]

A diana propicia foi Urbica, a santa priscilianista, a Gala aquitana, que expresaba o arroubo espiritual da nova relixión e o trasporte místico ante as atónitas campesiñas. [CORGA: M^a. X. Queizán, *Amantia*, Xerais, 1984]

Fisterra amosa a perfeta harmonía entre o mar e a terra [CORGA: GH2005-02-06/51]

unha saia vermella, tamén moi axustada, comprimida, marcáballe esaxeradamente os caprichos da carne contra o corpo, debuxándolle fofamente michelíns inecesarios que lle facían aparentar máis idade da que realmente tiña; [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

Na atualidade non usan a retórica de tempos pasados. [CORGA: CG2000-11-28/41]

Sulíñase tamén que a fixación de prezos, feita ata o de agora co amparo das institucións, restrinxía a libre competencia co cal fomentaba unha estrutura de mercado imperfecta. [CORGAetq: P1997-04/7]

É necesario que o esforzo lexislativo da aprobación dunha lei vexase acompañado dun incremento sustancial dos fondos que se adican ás actividades que se pretenden regular coa lei. [CORGAetq: ANT2002-10-31/5]

E tal derroche de xenerosidade serviu para que só no Ano Santo se realizasen 400 actividades _congresos, festas, encontros,..._ baixo o paraugas das conmemoracións do Xacobeo, algunhas tan trascendentes para Galicia como a Reunión de Científicos Latinoamericanos sobre Química, o congreso da Asociación Bíblica Española ou a reunión anual da Sociedade Luso Española de Neuroloxía. [CORGAetq: TN1998-04/13]

Mais a redución tamén a pode provocar a vocalización do primeiro elemento do grupo culto:

Antón sorprendéuse moito cando me ollou entrar no cuarto, e púxose a arrenegar da vella e dos tresvaríos maternais que lle daban agora a cada intre, dicindo que mellor houbera feito mandándolle a el direitamente o diñeiro gastado na viaxe, esquecendo o feito de que se atopaba xa en perfeitás condicións para derretelo, e de que en Madrid abundaban as mozas de bandeira e con boa saúde; [CORGA: L. Diéguez, *A canción do vagamundo*, Edicións do Castro, 1989]

"no podrán de modo alguno ser socios, ni como protectores, ni como numerarios, los que no profesen y practiquen la religión católica, apostólica y romana", debendo dar exemplo de boa conducta e cumpri-lo preceuto pascual. [CORGA: X. Palomares Ibáñez e I. Varela González, *Condición obreira, conflictividade e asociacionismo en Santiago na época da Comisión de Reformas Sociais*, Xerais, 1990]

Cecais a débeda histórica de comunidades tradicionalmente deficitárias, como a galega, poida calcularse así como a diferenza entre o custe efectivo e os recursos necesarios para soste un nivel de servizos aceitábel, en áreas como a educativa ou a sanitaria, desde as transferencias das mesmas ata hoxe. [CORGAetq: *GI1996-05/55*]

A aparición nos textos de formas coma as anteriores causa que na etiquetaxe manual xurdan, en paralelo coas correspondentes análises normativas pertencentes ás unidades que manteñen o grupo consonántico, moito máis habituais, as infrecuentes

trasporte	Scms	trasporte
perfeta	A0fs	perfeto
inecesarios	A0mp	inecesario
atualidade	Scfs	atualidade
suliña	Vpi30s	suliñar
sustancial	A0ms	sustancial
trascendentes	A0fp	trascendente
direitamente	Wn	direitamente
perfeitais	A0fp	perfeito
preceuto	Scms	preceuto
aceitábel	A0ms	aceitábel

4.3.2.2.4. Flutuacións X/S

Ás variacións aceptadas pola norma vixente do galego (VOLGa, 2004) e documentadas en moi diversas obras do corpus, como se mostra por exemplo nas unidades subliñadas das seguintes secuencias

Este exaxera a feminidade da súa personaxe de igual xeito que Harlow, Marilyn Monroe, María Montez ou Hedy Lamarr exaxeraban a súa. [CORGA: A. Pagán, *A mirada impasíbel. As películas de Andy Warhol*, Edicións Positivas, 2007]

Como facía Picasso, aínda que esaxeremos. [CORGAetq: *ANT2005-03-03/18*]

Nunha roda de prensa celebrada no Círculo de Belas Artes de Madrid, os convocantes da marcha invitaron á sociedade a secundar a manifestación e saír á rúa para exixir causas "tan xustas" como son o "si á paz" e o "non á guerra e ao terrorismo". [CORGAetq: GH2007-03-17/61]

Ademais, sinalou a "vocación de servizo" de todos eles, xa que "quen se integre na nosa candidatura sabe que se esixe unha dedicación absoluta", dixo. [CORGAetq: GH2007-03-30/21]

temos que engadir outras, para as cales a norma rexeita unha das dúas, malia que aínda así aparecen empregadas nos textos:

Aínda que semelle extraño é moito o que se ignora sobre o nazismo e de rebote poderíase dicir que é moito tamén o que se ignora sobre o nazismo propio, é dicir o franquismo. [CORGAetq: ANT2005-03-03/51]

A semana pasada tiven a ocasión de visitar San José e Miami cunha sensación estraña. [CORGAetq: GH2007-03-17/84]

"Sen ánimo de exhaustividade débese citar a potenciación da imaxe de Galicia en España e no extranxeiro, a súa positiva repercusión na economía galega, en especial no sector servicios, ou os investimentos realizados en construcións e outras obras que quedan á disposición da comunidade para o seu uso futuro", explican os xestores da Sociedade na memoria anual incluída no informe do Consello. [CORGAetq: TN1998-04/12]

Para frear a especulación financeira nos mercados internacionais, os asinantes optan por introducir un imposto sobre transaccións en moeda estranxeira. [CORGAetq: TN1997-07/29]

Fontes da compañía española precisaron que o acordo alcanzado con Percassi, estensible ó resto das cadeas de Inditex (matriz de Zara), terá como primeiro resultado a apertura dunha tenda de Zara e outra de Oysho (especializada en lencería e roupa interior) en Milán no primeiro semestre do ano que vén. [CORGAetq: CG2001-10-25/1]

Como me dixeron nunha ocasión (desexo que fago extensible aos que neste mes de xullo celebraron algo importante nas súas vidas e aos que en agosto vivirán momentos maravillosos), oxalá que a chuvia da felicidade e do amor nos pille co paraguas roto nunha tarde de soños, salpicando de ledicia a todos os que están ao noso arredor. [CORGAetq: GH2008-08-05/95]

Unha vez máis, a solución practicada na etiquetaxe manual e proposta para a automática é darlle cabida á variedade real dos textos, por moito que iso implique o recoñecemento e existencia de dúas formas e dous lemas, en todos os sentidos a mesma unidade exceptuando a grafía á que a presenza ou ausencia do fonema oclusivo velar xordo fai variar:

exaxera	Vpi30s	exaxerar
esaxeremos	Vps10p	esaxerar
exixir	V0f000	exixir
esixe	Vpi30s	esixir
extraño	A0ms	extraño
estraña	A0fs	estraño
extranxeiro	Scms	extranxeiro
estranxeira	A0fs	estranxeiro
estensible	A0ms	estensible
extensible	A0ms	extensible

4.3.2.3. Vacilacións ~ variacións na acentuación

A maior parte das discrepancias entre lemas irmáns procede de irregularidades na acentuación. Estas poden sistematizarse nos seguintes tipos:

4.3.2.3.1. Acento diacrítico

É máis frecuente do que esperabamos a ocorrencia de formas que por adición ou omisión dun til diacrítico, consciente ou non, causan serias dificultades no recoñecemento dos textos. Como podemos imaxinar son formas moi comúns na nosa lingua, e a falta do acento ou o seu engadido impiden que o etiquetador automático as identifique e clasifique correctamente. Por suposto, na etiquetaxe manual procedemos a aplicar a etiqueta e lema que corresponde, sen alterar a forma. Vexamos algún exemplo de cada un dos tipos.

a) Ausencia de acento diacrítico

O incremento da presenza galega na feira é un proxecto que se espera pór en práctica na edición do ano que ven. [CORGAetq: CG2001-10-17/7]

Segregar un país por razóns de lingua e un disparate. [CORGAetq: ANT2008-07-10/15]

Fronte a isto está o combate certo de Mero e Mini, cuxas armas son esas cordas tímbricas de zanfona, violín e guitarra, e nese concerto restitúense os acentos mais auténticos da nosa ancestralidade e proclámase a denuncia puntual contra os agravios e as desigualdades sociais. [CORGAetq: GH2008-08-05/94]

O texto aprobado so fala de traballar pola "desaparición progresiva de símbolos" nos edificios públicos (tamén os crucifixos e a Biblia nas tomas de posesión), sen provocar divisións, e advírtele á Igrexa Católica da caducidade dos seus privilexios. [CORGAetq: ANT2008-07-10/74]

A emigración era non poucas veces unha sorte de estratexia familiar, un xeito de diversificación dos riscos, e un procedemento para reforzar ou complementar cos ingresos de fora os rendementos da explotación rexida pola compañía familiar galega. [CORGAetq: GH2007-03-30/87]

Os erros máis usuais na falta de acento diacrítico nos textos do corpus de adestramento danse nos seguintes *tokens*¹⁵⁷:

ven por *vén*: 14 casos

e por *é*: 10 casos

mais por *máis*: 9 casos

so por *só*: 5 casos

fora por *fóra*: 4 casos

da por *dá*: 2 casos

compre por *cómpre*: 2 casos

A ausencia do til provoca que o etiquetador asigne a eses *tokens* a etiqueta que se axeita á forma gráfica da esquerda, nunca a que lle correspondería á forma gráfica da dereita. É dicir, caracteriza

ven imperativo de *vir* [V0m20s *vir*] ou 3ª plural do presente indicativo de *ver* [Vpi30p *ver*]

e conxunción coordinante [Cc *e*] ou substantivo [Scms *e*]

mais conxunción coordinante [Cc *mais*] ou conxunción subordinante [Cs *mais*]

so: preposición [P *so*]

fora 1ª ou 3ª do antepretérito de indicativo de *ser* e *ir* [Vli10s *ser*], [Vli30s *ser*], [Vlia0s *ser*], [Vli10s *ir*], [Vli30s *ir*], [Vlia0s *ir*] ou 1ª ou 3ª do pretérito de subxuntivo de *ser* e *ir* [Ves10s *ser*], [Ves30s *ser*], [Vesa0s *ser*], [Ves10s *ir*], [Ves30s *ir*], [Vesa0s *ir*]

¹⁵⁷ Quizais os máis numerosos sexan os que afectan á preposición *a* amalgamada co artigo feminino, pero neses casos a ausencia de til resulta moi difícil de avaliar pola confluencia na expresión de dúas das formas máis frecuentes nos textos: a preposición e o artigo feminino.

Pola contra, nós na etiquetaxe manual caracterizamos coa análise que en realidade lles corresponde, de xeito que as formas subliñadas nas oracións que encabezan este apartado, despois de realizar as modificacións oportunas na etiqueta e no lema, quedan como segue:

```
<forma>ven</forma>
<etiqueta>Vpi3s0</etiqueta>
<lema>vir</lema>
```

```
<forma>e</forma>
<etiqueta>Vpi3s0</etiqueta>
<lema>ser</lema>
```

```
<forma>mais</forma>
<etiqueta>Wm</etiqueta>
<lema>mais</lema>
```

```
<forma>so</forma>
<etiqueta>Wn</etiqueta>
<lema>so</lema>
```

```
<forma>fora</forma>
<etiqueta>Wn</etiqueta>
<lema>fora</lema>
```

b) Engadido de til diacrítico

Nos textos que conforman o corpus de adestramento non é inusual atopar palabras que non deberían estar acentuadas segundo as regras do galego, entre elas monosílabos, pero que reciben o acento gráfico probablemente por influxo do castelán. Trátase sobre todo do adverbio de afirmación e o pronome reflexivo *si*, mais tamén hai casos do pronome persoal *el* e dalgún demostrativo:

E anque a policía intenta demostrar que se trata dun suicidio, él esta convencido de outra cosa...
[CORGAetq: GH2007-03-30/108]

No exercicio económico, sí. [CORGAetq: GI1996-06/30]

En palabras dun importante operador turístico europeo, os ataques terroristas do 11 de setembro apuntillaron un ano xa de por sí nefasto, debido á recesión que orixinou o estourido da Intifada (levantamento) dos palestinos contra a ocupación israelí, o 28 de setembro do 2000. [CORGAetq: CG2001-10-04/49]

O mundo financeiro angloamericano sabe que ésta é a situación, e por tanto veremos unha aventura militar en Irak moi pronto. [CORGAetq: ANT2002-09-12/34]

Nos tres primeiros casos na etiquetaxe manual asignamos a etiqueta que lles corresponde e mantemos en forma e lema o til:

él	Rtn3ms	él
sí	Wn	sí
sí	Rtp3aa	sí

o último deles, porén, considerámolo unha variante en distribución complementaria –as formas do demostrativo en xeral acentúanse erroneamente cando actúan como non determinantes, namentres que cando funcionan como determinantes prescínlese do til–, polo que mantemos a forma inalterada e remitímola ao lema tipo:

ésta	Enfs	este
------	------	------

Non obstante, a relación de formas acentuadas innecesariamente cun suposto til diferenciador aumenta considerablemente se temos en conta todos os textos que prevemos etiquetar e non só os xa etiquetados. Neles, cunha frecuencia de aparición nada desdeñable, preséntanse por exemplo os seguintes monosílabos con falso til diacrítico:

<i>tí:</i>	563 casos
<i>dí</i>	318 “
<i>mín</i>	268 “
<i>pó</i>	130 “
<i>lá</i>	102 “
<i>dé</i>	38 “
<i>pá</i>	31 “
<i>rí</i>	12 “

Unha vez máis debemos sinalar que non son só os textos anteriores a 1982 os que presentan estas características, senón que tamén se recollen nalgúns publicados en pleno século XXI, como mostran as secuencias seguintes:

Mais, aínda que non coñeza tan ben coma tí as sabias palabras de Odín, sei que tamén dixo, pouco máis ou menos, rí e dille finuras a aquel no que non confies; [CORGA: X. Bernárdez Vilar, *A saga da illa sen noite*, Toxosoutos, 2001]

_Vaia o un polo outro _é o que dí. [CORGA: *ANT2003-01-16/75*]

Teño para mín que isto enlaza coa reclamación que xa fan algúns pais de poder leva-lo neno a almorzar e recollelo pola tarde non antes das seis. [CORGA: GH2007-03-25/47]

Mentres eles conversan a muller saca unhas agullas e un novelo de lá do bolso e ponse a facer a calceta... [CORGA: G. Pernas Cora, *Final de película*, Deputación Provincial da Coruña, 2006]

Afortunada e loxicamente, os números mostran un descenso acusado na pirámide cronolóxica a partir da implantación das *Normas* (RAG/ILG, 1982, 1995), mais tamén é de destacar o corte que se produce a partir do ano 2004, na nosa opinión debido a unha maior formación en galego pero, máis que nada, ao uso xeneralizado de ferramentas de corrección automáticas. Fixémonos senón nos datos que nos ofrecen as frecuencias de *tí*, extrapolables a calquera dos outros monosílabos:

Estadísticas por medio	Estadísticas por área temática	Estadísticas por lustro
Xornais <u>2</u>	Economía e Política <u>5</u>	1975-1979 <u>227</u>
Revistas <u>14</u>	Cultura e Artes <u>16</u>	1980-1984 <u>160</u>
Libros <u>547</u>	Ciencias Sociais <u>16</u>	1985-1989 <u>145</u>
	Ciencias e Tecnoloxía <u>1</u>	1990-1994 <u>11</u>
	Ficción <u>538</u>	1995-1999 <u>9</u>
	Outros <u>4</u>	2000-2004 <u>10</u>
		2005-2009 <u>1</u>
		Total: 563

4.3.2.3.2. *Adverbios en -mente*

O feito de que a acentuación dos adverbios en *-mente* siga regras distintas en galego e castelán provoca que, por interferencia, nos textos do CORGA se rexistren abondosos casos acentuados; tanto é así, que dos preto de tres mil cincocentos adverbios en *-mente* rexistrados na versión 1.5 do CORGA algo máis de trescentos figuran con til gráfico, o que representa un 8,57% das ocorrencias.

As cifras anteriores confirmannos que é unha desviación frecuente da norma, e por iso mesmo na etiquetaxe respectamos a diverxencia mantendo en forma e lema o *token* tal e como nos aparece nos textos, co que nos atopamos con unidades gráficas e lemas paralelos segundo presenten ou non acento:

Hoxe en día o ocio e o turismo rural é unha realidade difícil de cuantificar economicamente".
[ANT2005-03-03/58]

Segundo a investigación, o aproveitamento económico do monte "serviría para fixar poboación no rural, para desenvolver económicamente as zonas do interior e para contribuír ao relevo xeneracional". [CORGAetq: ANT2005-03-03/56]

economicamente	Wn	economicamente
económicamente	Wn	económicamente

Algúns outros pares dos que localizamos nos textos son os seguintes:

alfabeticamente	fraxilmente	habilmente
alfabéticamente	fráxilmente	hábilmente
criticamente	freneticamente	hipocritamente
críticamente	frenéticamente	hipócritamente
debilmente	friamente	histrionicamente
débilmente	fríamente	histriónicamente
cruamente	fuxidamente	identicamente
crúamente	fuxidamente	idénticamente
fisioloxicamente	graficamente	languidamente
fisiolóxicamente	gráficamente	lánguidamente

O manter esta etiquetaxe non supón ningún prexuízo para o etiquetador automático pois, a pesar de non figurar o lema nos lexicóns cos que traballa, categoriza ben os adverbios en *-mente* acentuados, salvo por suposto á hora de asignar lema. Contrástese, senón, o resultado da súa execución sobre as dúas seguintes secuencias inventadas:

Oración:	Pensándoo fríamente é boa idea.						
Formas:	pensando	o	fríamente	é	boa	idea	.
Etiquetas:	V0x000	Raa3ms	Wn	Vpi30s	A0fs	Scfs	Q.
Lemas:	pensar	o	*	ser	bo	idea	.

Oración:	Pensándoo friamente é boa idea.						
Formas:	pensando	o	friamente	é	boa	idea	.
Etiquetas:	V0x000	Raa3ms	Wn	Vpi30s	A0fs	Scfs	Q.
Lemas:	pensar	o	friamente	ser	bo	idea	.

Naturalmente, somos conscientes de que cando a variación gráfica non forme parte do estudo o feito de que esta se reflecta na etiquetaxe vailles causar máis dun problema aos usuarios do sistema; non obstante, esta pexa solventarase en versións posteriores do sistema coa inclusión do hiperlema, de xeito que poida decidirse nas consultas se interesan ou non os posibles resultados mantendo a variación ou ignorándoa.

4.3.2.3.3. *Acentuación nas voces graves rematadas en ditongo crecente*

É tamén recorrente nos textos que conforman o CORGA a presenza de unidades ortográficas graves rematadas en ditongo crecente que, por influxo do portugués ou porque se busca a diferenciación co castelán, reciben acento gráfico cando as normas do galego non o prescriben. Loxicamente esta desviación da norma percíbese xa no corpus de adestramento, aínda que case todos os casos se rexistran só en dúas noticias, unha anterior á primeira normativa oficial (*Man Común. Revista galega mensual de información xeral*, 1980) e a outra (*Galicia Internacional*, 1996) posterior á revisión das Normas de 1995 (RAG/ILG, 1995).

Algúns dos lemas documentados no primeiro dos textos reseñados son os seguintes: *burocracia, agrario, circunstancia, ciencia, coincidencia, competencia, comercio, consecuencia, contradictorio, dependencia, dominio, equilibrio, importancia, independencia, industria, inicio, materia, precio, presencia, servicio, socio*, etc.

De todas as voces que comprenden esta relación as máis comúns son no corpus de adestramento *necesario* e *proprio* con, respectivamente, dez e dezaseis ocorrencias, todas elas nos dous documentos sinalados máis arriba. Vexamos un exemplo aclaratorio:

Cecais a débeda histórica de comunidades tradicionalmente deficitarias, como a galega, poida calcularse así como a diferenza entre o custe efectivo e os recursos necesarios para soste un nivel de servizos aceptábel, en áreas como a educativa ou a sanitaria, desde as transferencias das mesmas ata hoxe. [CORGAetq: GII1996-05/55]

Agora ben, aínda que a súa frecuencia de aparición apenas supera a dunha ocorrencia, tampouco é imposible atopar exemplos noutros documentos posteriores a 1996. Unha mostra constitúena os seguintes:

Un discurso que nos amosa dous tipos de vida distintos: a dos habitantes de Abades, que ven a súa vida cotiá condicionada polas contínuas accións represoras, e a dos fuxidos, refuxiados en covas de difícil acceso, unha vida que dificilmente se pode chamar así. [CORGAetq: ANT2005-03-03/38]

"Explicálle aos da rúa o que viches aquí" a frase coa que se despide dela unha das reclusas bátelle na cabeza ate que uns trinta anos despois pon por escrito algúns dos episodios máis crueis da represión pero tamén historias de poderosa solidariedade carcelária. [CORGAetq: ANT2005-03-03/48]

A campaña de promoción de El Mundo e o partido de Rosa Díez, a conta da suposta "persecución do castellano", está retratando os prexuícios duns e a intención deliberadamente antinacionalista doutros. [CORGAetq: ANT2008-07-10/67]

Na etiquetaxe manual levada a cabo no corpus de adestramento mantense a variación amosada nos textos orixinais, de xeito que é común visualizar dúas formas e dous lemas distintos segundo presenten ou non acento gráfico:

preséncia	Scfs	preséncia	[5 casos no CORGA etiquetado]
presencia	Scfs	presencia	[55 casos no CORGA etiquetado]

Neste caso en concreto, de interesármonos polas construcións deste substantivo, teríamos que engadir aínda ás posibles variantes a forma que prescribe a normativa vixente:

presenza	Scfs	presenza	[49 casos no CORGA etiquetado]
----------	------	----------	--------------------------------

Unha vez máis, de querer empregar o corpus para estudar por exemplo a sufixación, a variación rexistrada tanto en formas como en lemas convértese nun atranco; con todo, o non respectala impediría os estudos que tivesen como obxecto, precisamente, a multiplicidade de opcións gráficas. Dado que o período temporal dos textos que pretendemos etiquetar abrangue dende 1975 ata a actualidade, cremos que non debemos prescindir na etiquetaxe e lematización da diversidade ortográfica, pois é unha circunstancia inherente á normalización de calquera lingua, incluído o galego. A inclusión do hiperlema nun futuro próximo solventará os inconvenientes actuais.

O etiquetador automático ante unidades que non figuran nos lexicóns cos que traballa nin no corpus de adestramento, caso de *desperdicios* e *adverténcia* presentes na secuencia de proba sobre a que o executamos e que amosamos a continuación, actúa coma se se tratase de unidades descoñecidas, e etiquétaas a partir das terminacións por

similitude con outras formas ortográficas semellantes rexistradas no lexicón ou no corpus e mais o contexto.

Oración:	Os desperdicios son poucos porque a adverténcia chegou cedo.									
Formas:	os	desperdicios	son	poucos	porque	a	adverténcia	chegou	cedo	.
Etiquetas:	Ddmp	Scmp	Vpi30p	Inmp	Cs	Ddfs	Scfs	Vei30s	Wn	Q.
Lemas:	o	*	ser	pouco	porque	o	*	chegar	cedo	.

4.3.2.3.4. *Acentuación nas voces agudas rematadas en ditongo decrecente*

Por influxo do castelán é moi frecuente atopar nos textos publicados con anterioridade a 1982 palabras agudas rematadas en ditongo decrecente acentuadas graficamente, pois nesta lingua as palabras que rematan en ditongos ortográficos acentúanse segundo as regras xerais de acentuación (*bonsái, habláis, estéis...*), regra que non se estende ao galego, xa que as normas desta lingua prescriben a ausencia de til gráfico nas voces agudas cando estas finalizan en ditongo decrecente.

Malia que sinalamos a data de 1982 por coincidir coa publicación da primeira edición oficial das normas ortográficas do galego e, en consecuencia, por percibirse claramente nas estatísticas un corte entre o antes e o despois, iso non implica que a partir dese ano xa non se documenten casos no corpus. Ben é verdade que a frecuencia de aparición nos dous últimos dos lustros que ofrece o CORGA (2000-2004 e 2005-2009) é residual, mais seguen aparecendo os tiles nalgúns nomes propios ou en voces de autores contados que eluden guiarse polas normas do galego:

"Pola súa especial interacción co cerebro, cada lingua foi modelando a través dos séculos o pensamento do seu pobo e posibilitando a recreación de mundos funcionáis" e "cada vez que unha lingua esmorece (...) é o tesouro universal quen se desfái de algo do que non se debera desprender" (ECG. 12/X/01). [CORGA: CG2001-10-17/36]

Ademais dos ataques aéreos contra os extremistas islámicos, o exército ruso volveu empregar foguetes, que nalgúns distritos, como no de Shatói, non utilizaba desde primeiros de ano. [CORGA: CG2000-11-28/69]

Por aquí pasóu todo o mundo. [CORGA: GH2005-02-06/65]

En Madrid, onde paséi eses días de aí atrás, soupen da venturosa noticia do acto de entrega dos galardóns, ó que me suméi mentalmente con fraternal afecto.. [CORGA: CG2000-11-25/38]

A bancada do Santiago Bernabéu chora de rabia cada vez que vai ver o Madrid de Fabio Capello. [CORGA: ANT2007-02-22/61]

Se consultamos a versión 1.5 do CORGA e extraemos os datos referentes ás formas acentuadas que rematan nalgún dos ditongos decrecentes tónicos, atopámonos coa seguinte distribución:

	-óu	-éu	-éi	-ái	-ói	-áu
1975-1979	2131	539	534	52	13	6
1980-1984	696	293	126	7	1	0
1985-1989	120	29	46	2	3	1
1990-1994	11	46	1	0	2	0
1995-1999	20	43	15	16	0	0
2000-2004	77	40	35	4	3	0
2005-2009	6	18	1	0	0	0
Total de casos	3061	1008	758	81	22	7

Os ditongos *óu*, *éu* e *éi*, son, por esa orde, os máis recorrentes; con todo, non debe estrañarnos a súa frecuencia de uso se pensamos en que unha das formas verbais máis utilizada nos textos, a terceira do singular, remata en *-ou* no pretérito dos verbos da primeira conxugación e en *-eu* nos da segunda; así mesmo outra das persoas moi empregadas en narrativa, a primeira de singular, acaba en *-ei* tanto no pretérito da clase máis numerosa como no futuro das tres conxugacións.

Se nos fixamos nas cifras indicadas na táboa, chamarannos moito a atención as dos dous últimos períodos, pois poderemos pensar que a norma de acentuación nos ditongos, lonxe de asentarse, segue interferida polo contaxio coa castelá; mais sería unha hipótese errónea, dado que ao achegármonos aos casos comprobaríamos que:

- Polo que respecta a *-óu*, os 6 casos utilizados entre o 2005 e o 2009 pertencen ao suplemento ou a artigos de opinión (*revelóu*, *tardóu*, *pasóu*, *ensinóu*, *empurróu*, *rubricóu*); das 77 ocorrencias de *-óu* atopadas entre o 2000 e o 2004, 62 localízanse en artigos de opinión do finado Agustín Sixto Seco no xornal *O Correo Galego*. Finalmente, os 20 resultados do lustro 1995-1999 repártense fundamentalmente en textos de Sixto Seco (*Caróu*, *inzóu*, *extrañóu*, *ficóu*, *estóu*), Camilo Gonsar (*adicóu*, *agregóu*, *chegóu*, *marchóu*, *regresóu*) e Miguel Ángel Rodríguez (*denuncióu*, *manifestóu*, *celebróu*, *aproveitóu*, *detectóu*).

- Polo que respecta ao ditongo *-éu*, no período 2005 a 2009, 16 das 18 ocorrencias totais correspóndense co antropónimo-topónimo *Bernabéu*. Entre o 2000 e 2004, das 40 aparicións de *-éu*, 19 pertencen ao nome do estadio madrileño, o *Santiago Bernabéu*, e 16 emprégaas Sixto Seco en artigos de opinión; algo semellante ocorre cos 43 casos do lustro 1995-1999, onde 37 casos son para o nome do estadio, 1 emprégao Sixto (*nacéu*) e os outros 5 Camilo Gonsar en *Arredor do non: caéu, doéu* (2), *erguéu, entendéu*.
- No referente a *-éi*, a única ocorrencia do último lustro xorde nun título (*O Señor Afranio ou como me rispéi das gadoupas da morte: memorias dun fuxido*); das 35 do 2000-2004, 23 débense a Sixto, 10 a propios –*Serguéi* (6), *Alexéi* (2), *Adréi* (1) e *Andréi* (1)–, 1 a Borobó (*agallopéi*) e outro máis a Bieito Iglesias (*encontréi*); por último, das 15 do período 1995-1999, 10 danse na adaptación do patronímico *Serguéi*, 3 emprégaas Gonsar en *Arredor do non (comentéi, probaréi, optéi)*, 1 recóllea Sixto (*lembréi*) e a que queda aparece na novela *A da alba sería (arelaréi)*.
- No concernente a *-ái* ocorre algo similar ca cos ditongos anteriores, só que a menor escala. No último dos lustros non se recolle ningún caso e no penúltimo só 4 –*Nikolái* (2), *Altái* e *desfái*– mentres que entre o ano 95 e o 99 documéntanse 15 ocorrencias, das cales 4 corresponden ao apelido *Shajrái*, 4 ao nome *Nikolái*, 7 ao substantivo *bonsái* (localizados todos na mesma noticia: *GI1995-09/18*) e a última delas é un erro de lectura do OCR (*mái* por *máis*, que aparecerá corrixido na próxima versión do CORGA).
- Os tres casos de *-ói* do lustro 1990-1994 apenas merecen atención, pois volvemos atoparnos con topónimos, dous estranxeiros (*Shatói* e *Achjói*) e un local (*Raxói*). As demais ocorrencias do corpus reflicten sobre todo nomes propios (*Donskói, Elói, Tolstói, Calói*) e a forma verbal *fói*.
- Finalmente, no tocante ao ditongo *-áu*, o menos utilizado por non coincidir con terminación verbal nin darse en unidades frecuentes nos textos, recollemos no CORGA só 7 ocorrencias, documentadas ademais en só dúas obras; 6 delas pertencen ao substantivo *bacalláu* e localízanse nas novelas de

Fernández Ferreiro, *Morrer en Castrelo do Miño* (5 casos), e Fariña Jamardo, *Golfaróns de sangue* (1 caso), mentres que a sétima é *ribeiráu* e aparece novamente en Fernández Ferreiro.

No corpus de adestramento non atopamos ningún caso dos anteriores; mais, de apareceren, seguiríase con eles a mesma práctica que en circunstancias semellantes: etiquetaxe da forma tal e como se documenta. Exs.:

O personaxe con máis posibilidades para encabezar un goberno de transición en Grozni, capital de Chechenia, sería en principio o novo virrei desa república, Nicolái Semiónov, que desprazou ó ata agora home forte en Grozni do Kremlin, Nicolái Yegórov, afectado por problemas de saúde. [CORGA: CG1995-02-03/63]

Nicolái Semiónov	Spm0	Nicolái Semiónov
Nicolái Yegórov	Spm0	Nicolái Yegórov

Debéu ser eso o que tamén enrabechóu ó enxeñeiro do que falo, que interpretóu o acordo como un agravio a toda unha comarca. [CORGA: CG2000-11-01/31]

debéu	Vei30s	deber
enrabechóu	Vei30s	enrabechar
interpretóu	Vei30s	interpretar

No recoñecemento automático, en xeral, a caracterización morfolóxica é de novo realizada correctamente, como podemos comprobar na seguinte secuencia coa que testamos o etiquetador, malia que este non asigne lema por tratarse de formas non presentes no lexicón (*asubiéi, cantóu*):

Oración:	Eu aínda non asubiéi e o paxaro xa cantóu.									
Formas:	eu	aínda	non	asubiéi	e	o	paxaro	xa	cantóu	.
Etiquetas:	Rtn1as	Wn	Wn	Vei10s	Cc	Ddms	Scms	Wn	Vei30s	Q.
Lemas:	eu	aínda	non	*	e	o	paxaro	xa	*	.

4.3.2.3.5. Acentuación nos hiatos

Finalmente, para rematar coas características acentuais máis comúns dos textos que pretendemos etiquetar, quedáanos só por comentar a ausencia de til sobre a vogal pechada tónica, *i* ou *u*, cando se encontra en hiato.

Neste punto constátase nos textos galegos unha interferencia por partida dobre; por unha banda, mestúrase a regra galega de acentuar graficamente as vogais *i*, *u* tónicas cando forman sílaba (RAG/ILG, 1982: 17; 1995: 25; 2003: 27-28) coa consideración en castelán de ditongos a efectos ortográficos da combinación das dúas vogais cerradas distintas (RAE, 1999: 43; RAE, 2010: 235) (*diurno*, *viuda*, *ruido*, etc.), independentemente da súa posible articulación como hiatos. Pola outra banda, interfere a regra de acentuación do portugués (Cunha e Cintra, 2005: 71) segundo a cal non se acentúan o *i* e o *u* tónicos cando ocupan a coda da sílaba as consoantes *l*, *m*, *n*, *r*, *z* ou o dígrafo *nh* (*adail*, *contribuinte*, *demiurgo*, *juiz*, *paul*, *retribuirdes*, *ruim*, *ventoinha*, etc.).

Unha mostra da interferencia castelá ou quizais só o descoñecemento da norma galega no corpus de adestramento constitúena as tres secuencias seguintes:

Prexuízos europeos [CORGAetq: TN1997-04/7]

Pouca xente, poucos actores, que dan voltas ao redor dun chinés viuvo e xubilado, que chega a unha cultura diametralmente distinta, coa mente aberta e disposto a falar con calquera, incluídos os mormóns (unha das escenas máis felices do filme) e a aprender as novidades que lle achega unha cultura que non pode entender de todo. [CORGAetq: ANT2008-07-10/44]

Ir lendo como todo o machaca e despois como lle vai collendo interese a vivir, a partir dun feito fortuito. [CORGAetq: ANT2008-07-10/29]

Á parte dos substantivos e adxectivos que coma *xuízo*, *gratuíto*, *diúrno*, etc. posúen un *i* ou *u* tónicos en hiato e deixando fóra os posesivos femininos *súa(s)* (3219 casos) e *túa(s)* (307 ocorrencias), a maior parte dos erros de acentuación neste sentido no CORGA danse nos verbos rematados en *-uír* (*afluír*, *atribuír*, *concluír*, *construír*, etc.), e en menor medida nos de *-oír* (fundamentalmente, *oír*) e *-aír* (sobre todo *sair* e *caír*). Tanto é así que constatamos no CORGA unha frecuencia de 1 contra 6 no uso do infinitivo dos verbos finalizados en *-buír* fronte aos de *-buír*¹⁵⁸:

¹⁵⁸ Ao non estar etiquetado é imposible realizar buscas por lemas ou incluso distinguir, como non sexa á man, entre a secuencia *iu*, *ui* cando son ditongos de cando son hiatos e non están acentuados ou mesmo cando na secuencia *ui*, o *u* forma parte do dígrafo *gu* (*seguir*, *seguíu*, *seguido*...). De aí que restrinxiramos a consulta aos verbos rematados en *-buír* en vez de estendela aos acabados en *-uír*.

	Estadísticas por medio	Estadísticas por área temática	Estadísticas por lustro
-BUIR	Xornais 20	Economía e Política 65	1975-1979 10
	Revistas 37	Cultura e Artes 26	1980-1984 14
	Libros 56	Ciencias Sociais 35	1985-1989 23
		Ciencias e Tecnoloxía 15	1990-1994 9
		Ficción 6	1995-1999 42
		Outros 16	2000-2004 12
		2005-2009 3	
		Total: 113	

	Estadísticas por medio	Estadísticas por área temática	Estadísticas por lustro
-BUÍR	Xornais 303	Economía e Política 294	1975-1979 2
	Revistas 76	Cultura e Artes 84	1980-1984 5
	Libros 234	Ciencias Sociais 158	1985-1989 23
		Ciencias e Tecnoloxía 100	1990-1994 86
		Ficción 88	1995-1999 182
		Outros 111	2000-2004 170
		2005-2009 145	
		Total: 613	

Naturalmente, se nos fixamos nas cifras, tamén comprobamos como nos textos se produce un incremento paulatino e progresivo da acentuación correcta e, en contrapartida, a case desaparición da práctica errónea.

No lexicón non introducimos unha conxugación paralela á normativa nun primeiro momento por entender que as formas que non seguisen a normativa oficial serían as menos e se documentarían en textos previos á redacción das normas oficiais; contabamos ademais coa hipótese de que non aparecerían nos textos recentes que pretendiamos etiquetar e converter en corpus de adestramento, polo que a súa inclusión no lexicón quedaría para bastante máis tarde. Non obstante, algúns dos textos de adestramento emperráronse en demostrar o desacerto da nosa suposición e obrigáronnos a lidar con unidades coma as seguintes:

atribuir	V0f000	atribuir
concluir	V0f000	concluir
constituída	V0p0fs	constituir
constituiron	Vei30p	constituir
construir	V0f000	construir
contribuir	V0f000	contribuir

destruir	V0f000	destruir
diminuir	V0f000	diminuir
diminuiu	Vei30s	diminuir
excluidas	V0p0fp	excluir
excluir	V0f000	excluir
excluíron	Vei30p	excluir
incluídos	V0p0mp	incluir
incluir	V0f000	incluir
influir	V0f000	influir
reconstituír	V0f000	reconstituír

Seguindo o criterio de respectar o orixinal e fomentando a coherencia co resto da etiquetaxe, decidimos manter as formas tal cal estaban e remítilas ao seu lema correspondente, excluindo deste o acento normativo. Nesta decisión primou a aparición na propia unidade ortográfica da secuencia *ui* co *i* tónico sen acentuar, polo que nos pareceu oportuno trasladala tamén así ao lema.

Agora ben, nos casos en que a unidade gráfica referencia unha forma verbal de verbos en *-uír* e esa mesma secuencia *-iu, ui-* é ditongo ou non aparece, optamos por remitir a unidade ao lema normativo, pois a grafía da forma non nos aconsella outra solución. Exemplos do corpus de adestramento son:

contribuiría	Vci30s	contribuír
constitue	Vpi30s	constituír
construírá	Vfi30s	construír
contribuiremos	Vfi30p	contribuír
destruírá	Vfi30s	destruír

Na etiquetaxe automática, o tratamento destas unidades dependerá de se se rexistrou algún caso no corpus de adestramento, pois xa adiantamos que non figuran no lexicón. Se a resposta é afirmativa, vai etiquetalas correctamente e, se é negativa, dependerá de múltiples factores, xa que o etiquetador actuará coma se se tratase de palabras descoñecidas e terá en conta os pares forma-etiqueta anteriores e posteriores, a terminación da unidade, etc.

Executamos o etiquetador automático sobre tres oracións que conteñen unha unidade semellante ás que estamos tratando e que non aparecen no corpus de adestramento: *retribuir*, *construido* e *construen*. Nos dous primeiros casos é probable que o acerto veña determinado pola terminación e a frecuencia das perífrases nas que incluímos eses *tokens* no corpus de adestramento; o erro na adscrición do modo da terceira, en cambio, non sabemos explicalo, posto que a terminación *-en* aparece 1335 veces asociada ao indicativo e só 236 ao subxuntivo:

Oración:	Hai que retribuir máis o bo traballo.							
Formas:	hai	que	retribuir	máis	o	bo	traballo	.
Etiquetas:	Vpi30s	Cs	V0f000	Wn	Ddms	A0ms	Scms	Q.
Lemas:	haber	que	*	máis	o	bo	traballo	.

Oración:	Non dei construido unha torre con estes bloques.								
Formas:	non	dei	construido	unha	torre	con	estes	bloques	.
Etiquetas:	Wn	Vei10s	V0p0ms	Difs	Scfs	P	Edmp	Scmp	Q.
Lemas:	non	dar	*	un	torre	con	este	bloque	.

Oración:	Os nenos construen torres con caixas.						
Formas:	os	nenos	construen	torres	con	caixas	.
Etiquetas:	Ddmp	Scmp	Vps30p	Scfp	P	Scfp	Q.
Lemas:	o	nenos	*	torre	con	caixa	.

4.3.2.3.6. Acentuación esdrúxula na P4 e P5 do copretérito, pospretérito e antepretérito de indicativo e mais o pretérito de subxuntivo

Nos documentos do CORGA coexisten, igual ca na fala, as acentuacións paroxítonas e as proparoxítonas na primeira e segunda persoas do plural dos tempos copretérito, pospretérito e antepretérito de indicativo e pretérito de subxuntivo. Malia que neste último a norma vixente dende o 2003 do galego prescribe a acentuación esdrúxula, non ocorre así antes desa data nin co resto de tempos, para os que a acentuación grave representa a orixinaria en tanto a proparoxítona é produto da interferencia co castelán.

Vexamos algúns exemplos reais:

Hai poucos anos estábamos no 0,03%, e no 2003 chegamos ao 0,06%. [CORGAetq: ANT2002-10-31/5]

¿En canto poderíamos cifrar o potencial promocional do programa? [CORGAetq: GI1996-06/34]

Por moitas veces que o intentáramos, o resultado sería sempre o mesmo: [CORGA: M. gondar Portasany, Prólogo a *A violencia excedente*, Laiovento, 2006]

Pois logo sabede que para todo isto é para o que están as editoriais discográficas, aquelas das que comezáramos a falar. [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da ría*, Xerais, 2005]

Preferían pagarlle unha habitación e que o aturásemos nós. [CORGA: B. Iglesias, *A historia escríbese de noite*, Galaxia, 2001]

Coherentemente, na etiquetaxe manual realízanse as análises correspondentes a estas formas coa acentuación proparoxítona:

estábamos	Vii10p	estar
poderíamos	Vci10p	poder
intentáramos	Ves10p	intentar
comezáramos	Vli10p	comezar
aturásemos	Ves10p	aturar

Na etiquetaxe automática, con todo, prodúcese unha desfeita no recoñecemento de formas coma as anteriores, visible nas dúas secuencias seguintes coas que testamos o funcionamento do etiquetador automático, feito este que nos aconsella integrar tamén no lexicón a acentuación proparoxítona para a P4 e P5 dos tempos citados máis arriba:

Oración:	Estábades vós no enterro?					
Formas:	estábades	vós	en	o	enterro	?
Etiquetas:	Sp00	Rtn2ap	P	Ddms	Sems	Q?
Lemas:	*	vós	en	o	enterro	?

Oración:	Se quixeramos, partíramolo nós.							
Formas:	se	quixera	me	os	,	partíramolo	nós	.
Etiquetas:	Cs	Ves30s	Rad1as	Raa3mp	Q,	Scms	Rtp1ap	Q.
Lemas:	se	querer	me	o	,	*	nós	.

Finalmente, gustaríanos subliñar que o emprego da acentuación esdrúxula non é sistemático nun mesmo autor nin se documenta só nos textos máis antigos do corpus. A vacilación na acentuación chega ata os nosos días:

Así teño que dicir que estivo mal que se nos catalogase como políticos ou que se dixera que as voces que defendiamos ao noso mar defendiamos políticas partidarias, cando, en realidade, só cumpríamos coa nosa obriga como patróns maiores de dicir o que estaba a pasar e non calábamos fronte ao poder. [CORGAetq: ANT2005-03-03/24]

"Grazas á globalización, a vitoria aliada na II Guerra Mundial e o liderado americano en ciencia e tecnoloxía, o inglés alcanzou un éxito no mundo que vai máis aló do que podíamos agardar. [CORGAetq: ANT2008-07-10/18]

4.3.2.4. Vacilacións ~ variacións na adscrición xenérica ou numérica

Malia que non repercuta en xeral na lematización, si son tamén abondosos os substantivos que, por interferencia co castelán, se empregan cun xénero que, normativamente, non lles corresponde en galego. Algúns exemplos son:

Despois tivo lugar a mesa redonda no Ateneo da capital ourensá, presentada polo humorista Xose Lois González O Carrabouxo, quen asegurou que se trata de homenaxe "necesario e bo", porque "aínda a calor da presenza de Peña na rúa do Paseo non se foi". [CORGAetq: GH2007-03-17/26]

Se queres sabelo, ofrézoche facer un viaxe comigo á Coruña. [CORGA: X.M. Martínez Oca, *Beiramar*, Xerais, 1983]

O calor desfixose en pedriñas de xeo dentro das catro paredes que xa nada contiñan. [CORGA: X. Cuentas Castro, *O pescador de cana*, Toxosoutos, 2006]

Atopounos na mañá ceda a Lourida, que baixaba dende Seixadela a vender a leite nas casas dos señoritos da vila, e que recoñeceu a meu pai. [CORGA: X. Fariña Jamardo, *Golfaróns de sangue*, Edicións do Castro, 1989]

Na etiquetaxe manual modificamos na etiqueta o valor do xénero do substantivo que figura no lexicón polo que aparece reflectido no documento. Así:

<i>Forma</i>	<i>Etiqueta no lexicón</i>	<i>Etiqueta na secuencia anterior</i>
homenaxe	Scfs	Scms
viaxe	Scfs	Scms
calor	Scfs	Scms
leite	Scms	Scfs

Desafortunadamente, na etiquetaxe automática predomina a caracterización do substantivo que figura no lexicón sobre o xénero do determinante que o precede ou o do adxectivo que o segue nunha secuencia calquera, de xeito que o contexto non se utiliza nestes casos para nada; así pois, non hai forma de realizar o cambio no valor a non ser que introduzamos no lexicón a posibilidade dos dous valores xenéricos, o que non nos parece correcto pois incrementaría proporcionalmente a ambigüidade. Non obstante, debemos pensar que estas estruturas resultarán fáciles de recuperar xa que, ao permitir a consulta simultánea de ata catro formas, lemas e/ou etiquetas, temos a opción de introducir antes o xénero do determinante oposto ao do substantivo, ou despois o do adxectivo, ou combinando ambos. Por exemplo:

Formal:	<input type="text"/>	Forma2:	<input type="text"/>	Forma3:	<input type="text"/>	Forma4:	<input type="text"/>
Etiquetal:	*dms	Etiqueta2:	Scfs	Etiqueta3:	<input type="text"/>	Etiqueta4:	<input type="text"/>
Lemal:	<input type="text"/>	Lema2:	<input type="text"/>	Lema3:	<input type="text"/>	Lema4:	<input type="text"/>
Unidade1:	<input type="text"/>	Unidade2:	<input type="text"/>	Unidade3:	<input type="text"/>	Unidade4:	<input type="text"/>
Área temática:	Todas <input type="button" value="v"/>			Subárea:	Todas <input type="button" value="v"/>		
Dende:	1975 <input type="button" value="v"/>	Ata:	2010 <input type="button" value="v"/>	Medio:	Todos <input type="button" value="v"/>		
	Corpo	Pé de foto	Nota	Titular	Resumo	Acoutación	
Buscar	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
en:	Prólogo	Apéndice	Dedicatoria	Cita	Encabezamento		
	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
							<input type="button" value="Limpar"/> <input type="button" value="Buscar"/>

Polo que respecta ao número, non é inusual tampouco atopar nos documentos do CORGA os típicos casos de substantivos que aínda designando un só obxecto, precisamente por estar este constituído por partes simétricas, se empregan en plural en

vez do singular normativo, co que alternan: *tesoiras* vs. *tesoira*; *pantalóns* vs. *pantalón*; *bragas* vs. *braga*, etc.

Algunhas veces tamén aceptaba algo, un xabón especial, unhas tesoiras, un peite, calquera obxecto que lle ofrecesen e non fose de moito custo. [CORGA: D.X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Calquera cousa que se interpoña é unha tesoira que ameaza ese fio. [CORGA: X. Méndez, *As horas que nos quedan*, Sotelo Blanco, 1997]

Outro tripulante, Manuel de Laureano, "O Neto", baixou os pantalóns e ensinoulle-lo cu como burla do desafío. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

E daquela Otegi e Zapatero había poucos anos que deixaran o pantalón curto. [CORGA: *ANT2007-03-15/5*]

O proceso contrario, a singularización, prodúcese en substantivos que normativamente son plurais e que é moi raro velos utilizados en singular; con todo, algún caso hai:

O outro abre a boca e amósalle unha fauce enorme, negra e profunda cuns chavellos grandes, amarelos e desiguais na entrada e cunhas grandes moas negras detrás. [CORGA: S. de Toro, *Tic tac*, A Nosa Terra, 1993]

Admireime que estivera só e comprendín que se berraba era para que o seu compañeiro se dera présa en xurdir, pois estaban facéndome a tenace. [CORGA: H. Puentes, *O bandido Casanova*, Xerais, 2000]

Os máis frecuentes son, como non, os plurais respectivos:

Non todos os cordeiros descarreirados rematan nas fauces do león. [CORGA: X.C. Fernández Caínzos, *Os espidos pés do destino*, Bahía, 2000]

Por un intre pensei que podía facelo, que ía en serio, e que aquelas tenaces de robot estaban a pique de afogarme contra a parede. [CORGA: M. Rivas, *Os comedores de patacas*, Xerais, 1993]

Na etiquetaxe do corpus de adestramento non apareceu aínda ningún caso destes, mais, de apareceren, serían etiquetados e lematizados en consecuencia, ben con singular ben con plural, dependendo sempre da ocorrencia no seu contexto. Por exemplo, a etiquetaxe correspondente aos pares subliñados das secuencias anteriores diferenciaríase no valor numérico da etiqueta e, por suposto, na lematización de cada integrante dos pares:

tesoiras	Scfp	tesoiras
tesoira	Scfs	tesoira
pantalóns	Scmp	pantalóns
pantalón	Scms	pantalón
fauce	Scfs	fauce
fauces	Scfp	fauces
tenace	Scfs	tenace
tenaces	Scfp	tenaces

É tamén moi habitual nos textos que imos etiquetar a aparición do pronome de dativo de terceira persoa, *lle*, con valor plural, circunstancia que traslada á escrita a non distinción nunha boa parte da xeografía galega entre o referente singular e o plural. Este fenómeno xa se recolle no corpus de adestramento, onde o atopamos en oposición a *lles*

Os estudos falan dun mozo de entre vinte e trinta anos, cun salario que oscila entre os 600 e 900 euros e que traballa máis horas das que lle corresponden por contrato. [CORGAetq: ANT2005-03-03/38]

O número de accións que Zeltia pensa venderlles como máximo ós beneficiarios do plan ascende a 300.000 accións, representativas do 0,1878 por cento do seu capital social. [CORGAetq: CG2000-11-30/8]

lle	Rad3ms	lle
[vender] <u>lles</u>	Rad3mp	lle

ou como único integrante do paradigma de terceira dativo:

Este centro, ademais de achegarlle formación ás persoas con discapacidade, terá tamén unha área de formación para profesionais e os medios necesarios para a utilización das novas tecnoloxías. [CORGAetq: GH2007-03-17/32]

Zapatero e Blanco déronlle persoalmente á benvinda aos convidados a través de videochamada, o gabinete de comunicación utilizou a nota de prensa virtual para convocar os xornalistas e o partido creou a figura do "ciber-delegado" para que calquera internauta, ademais de seguir en directo o congreso, puidera enviarlle mensaxes á dirección no transcurso dos debates. [CORGAetq: ANT2008-07-10/76]

O debate de onte na TVG, pese a ser moi agardado, apenas tivo picos de enfrontamento dos que lle gustan aos telespectadores. [CORGAetq: DLV2007-05-15/17]

Ou seu partido administra mal o resultado, non é dialogante, non é capaz de pactar, convertendo en inútil o voto de moita xente", advertiulle o portavoz do BNG ao líder da oposición. [CORGAetq: DLV2007-05-15/17]

[achegar] <u>lle</u>	Rad3fp	lle
[déron] <u>lle</u>	Rad3mp	lle
lle	Rad3mp	lle
[advertiu] <u>lle</u>	Rad3ms	lle

Na etiquetaxe automática a caracterización da unidade *lle* é, polo momento, sempre singular, pois ese é o valor que consta no lexicón. Agora ben, se queremos que o etiquetador caracterice tamén o valor plural que posúe en numerosos casos, debemos incorporar esa posibilidade na etiqueta que o defina no lexicón, de xeito que deben coexistir estas dúas entradas:

lexicón	sublexicón	raíz	subetiqueta	grupo	lema	categoría	normativa	fonte
xeral	principal	lle	Rad3a	G29	lle	R	si	volga 2004
xeral	principal	lle	Rad3ap		lle	R	si	volga 2004

A ampliación na caracterización deste pronome é de suma complexidade no sistema de XIADA, pois a modificación realizada no lexicón só repercute nos casos en que *lle* aparece como unidade simple, é dicir, entre espazos en branco, pero non lle afecta á súa aparición en contraccións ou en énclise. A inclusión do valor plural, por exemplo, nas contraccións implica a introdución das seguintes características:

contracción	normatividade	forma	etiqueta	lema		forma	etiqueta	lema
llo	n	lle	Rad3ap	lle	/	o	Raa3ms	o
llo	n	lle	Rad3fp	lle	/	o	Raa3ms	o
llo	n	lle	Rad3mp	lle	/	o	Raa3ms	o
llos	n	lle	Rad3ap	lle	/	os	Raa3mp	o
llos	n	lle	Rad3fp	lle	/	os	Raa3mp	o
llos	n	lle	Rad3mp	lle	/	os	Raa3mp	o
lla	n	lle	Rad3ap	lle	/	a	Raa3fs	o
lla	n	lle	Rad3fp	lle	/	a	Raa3fs	o
lla	n	lle	Rad3mp	lle	/	a	Raa3fs	o
llas	n	lle	Rad3ap	lle	/	as	Raa3fp	o
llas	n	lle	Rad3fp	lle	/	as	Raa3fp	o
llas	n	lle	Rad3mp	lle	/	as	Raa3fp	o

Máis complexo aínda é engadir o valor plural cando o pronome *lle* aparece nunha secuencia de clíticos ou enclítico a unha forma verbal, pois hai que redefinilo na táboa de clíticos monosilábicos e modificar algunhas das regras lingüísticas que rexen a súa caracterización morfosintáctica. Por exemplo, a regra na que se indicaba que se era

o último elemento da unidade só podía ser singular debemos eliminala pois xa non é acertada¹⁵⁹.

4.3.2.5. Vacilacións ~ variacións no caso acusativo e dativo

En xeral, a oposición entre acusativo e dativo está clara na xeneralidade do galego e así se manifesta nos textos do CORGA.

Deste xeito, na versión 1.5 do CORGA atopámonos con:

- verbos que só presentan a estrutura transitiva, coma *buscar* ou *insultar*:

Búscalo entre os discos que hai no moble da sala e sube ben o volume, a ver se a música desinfecta a casa. [CORGA: F. X. Fernández Naval, *Tempo de Crepúsculo*, Xerais, 1993]

O taxista endereitouse por fin sobre o volante, e esperou a que ela baixase do coche para insultala a través da xanela: [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

- verbos tales como *abusar* ou *caducar* que unicamente se empregan como intransitivos:

A filosofía da linguaxe no século XX abusou da poderosa metáfora conceptual da "lingua traizoeira". [CORGA: T. Moure, *Outro idioma é posible. Na procura dunha lingua para a humanidade*, Galaxia, 2005]

A calefacción era excesiva e a súa carta de identidade caducaba hoxe. [CORGA: X. Borrazás, *Contos malvados*, Xerais, 1998]

- verbos que segundo se constrúan con complemento directo ou indirecto posúen acepcións distintas, caso de *dormir* ou *servir*:

a) transitivos

dormir = facer prender no sono;

Acoutación: (O rei bótase enriba da cama e Rouco, despois de tapalo ben, dúrmeo ao son dunha nana). [CORGA: R. Vidal Bolaño, *Laudamuco, señor de ningures / Ledaiñas pola morte do Meco*, Pico Sacro, 1977]

servir = pórse a disposición

¹⁵⁹ Nas combinacións de clíticos e nas regras de clíticos aparece a expresión *lle* co valor de plural, mais a forma que realmente referencia aí é *lles*, que por rexistrarse nunha contracción coa segunda forma do artigo ou co pronome de acusativo terceira, graficamente perde o *s*.

Cando volví de Madrid viña pensando que se soubese para onde marchara, iría de cabeza detrás dela e ofreceríamelle para servila en todo o que cumprise, sen pedir nada a cambio, só estar a carón dela o resto da miña vida... [CORGA: X. M. Martínez Oca, *Náufragos en terra. Os amores imposibles*, Xerais, 1995]

b) intransitivos

durmir = pasar a noite para repousar

a primeira noite durmiremos en Zamora, despois Salamanca e Plasencia, desviámonos a Yuste e Guadalupe, pasamos por Mérida e durmimos en Llerena; [CORGA: X. A. Perozo, *Martázul, Ir Indo*, 2001]

servir = ser útil, apto ou axeitado

Luísa: Tes unha bolsa na que poderías meter as laranxas pero a bolsa está rota e non che serve... [CORGA: A. Lopo, *Os homes só contan até tres*, Xerais, 2005]

- verbos, finalmente, que flutúan entre a pronominalización con pronome de acusativo e a de pronome de dativo, tales coma *axudar*, *obrigar*, *informar*, etc.:

O asistente persoal acompañará e axudalle á persoa dependente para facilitarlle o seu acceso á educación, ao traballo e a outros ámbitos da vida. [CORGAetq: *GH2007-03-17/32*]

Bertelsmann, o conglomerado propietario da discográfica BMG, prestáralle a Napster 85 millóns de dólares para axudalo a construír un servizo de subscrición "legal". [CORGAetq: *CG2002-05-19/55*]

[axudara]lle	Rad3fs	lle
[axuda]lo	Raa3ms	o

Non obstante todo o anterior, no corpus de adestramento, e por extensión tamén no CORGA, atopámonos con verbos que, malia construírse normalmente na fala e tamén segundo a norma con acusativo, se empregan no noso corpus en ocasións cun pronome cuxa expresión é a de dativo. Trátase, sobre todo, dos verbos *levar*, *instar* e *advertir* nas seguintes acepcións:

levar: *v.t.* 8. prop.e fig. Ser [unha persoa ou cousa] o medio ou a causa que fai que [alguén] chegue [a certo sitio, dun sitio a outro, ou a certo resultado]. OBS. Úsase coa preposición *a*. *As súas investigacións levárono á solución do caso. O dictador levou o país á ruína.* [DRAG, 1997: 706]

instar: v.t. 1. Facer obxecto [a alguén] dunha insistente e firme petición ou solicitude. OBS. A petición vai introducida polas preposicións *a* e *para*. *Instouno para que resolvese canto antes o seu problema*. [DRAG, 1997: 665]

advertir: v.t. 2. Facer saber [algo que debe ser tomado en consideración]. OBS. En certas circunstancias vai seguido da prep. *de*. *O médico advertiume que non tomase alcohol durante o tratamento*. CF. **recomendar**. *Advertíronme que era perigoso asomarse. -¿Sabíades que era necesario o pasaporte? -Non, ninguén nolo advertiu*. CF. **dicir, informar**. *Advertíronnos das dificultades que iamos encontrar*. SIN. **avisar, previr**. *Os sinais advirten claramente da existencia de curvas perigosas*. [DRAG, 1997: 33]

Vexamos algúns dos exemplos:

Recordou que desde a antiga oficina principal, Caixa Galicia iniciou unha expansión pola cidade olívica que lle levou a contar na actualidade cun total de 25 oficinas que proximamente se verán aumentadas en número. [CORGAetq: CG1996-02-15/38] [lle Raa3fs lle]

O xogador serbio saltou ó Estadio de Cayo Biscaíno moi concentrado en non repetir os erros que lle levaron a perder en Indian Wells e na segunda ocasión de que dispuxo logrou romper o servizo do xogador español para colocarse cun 3-1 ó seu favor. [CORGAetq: GH2007-03-30/75] [lle Raa3ms lle]

Nadal intentou reaccionar, pero o gran xogo do seu rival levoulle a ceder a primeira manga por 6-3. [CORGAetq: GH2007-03-30/75] [lle Raa3ms lle]

Aínda que as reivindicacións duns e doutros son diferentes a suma de ambas as protestas afectou a numerosos pacientes, que onte se queixaban fundamentalmente de que ninguén lles informou da xornada de paro. [CORGAetq: GH2007-03-30/8] [lles Raa3mp lle]

A cantidade obxectada deberase ingresar nalgún proxecto alternativo (non se nega a capacidade recadatoria do Estado, só a recadación para fins militares), e o comprobante deste ingreso inclúese na documentación da declaración do IRPF xunto cunha carta dirixida ó Delegado da Facenda expresándolle os termos da obxección fiscal, e instándolle a que no futuro os impresos inclúan a liña da obxección fiscal para recoñecer e facilitar este dereito. [CORGAetq: TN1997-06/7] [lle Raa3ms lle]

o axente preguntoulle se posuía carné de conducir, e ó responderlle afirmativamente, advertiulle de que ía perder tres puntos. [CORGAetq: GH2007-03-17/40] [lle Raa3as lle]

O feito máis que probable de que detrás da maioría destes usos con pronome de dativo estea o castelán –o cal, debido á extensión entre falantes cultos e escritores de prestixio, admite o uso de *le* no sitio de *lo* en función de complemento directo cando o referente é unha persoa do sexo masculino (DPD, 2005: 263)–, unido ao rexeitamento que

como falantes e lingüistas nos producía a aparición desas formas, levounos nun primeiro momento a etiquetalas como acusativo, tal e como se amosa ao final de cada exemplo.

Un pouco máis complexos ca os casos de *levar* e *instar* mais pronome de dativo resultaban, ao noso modo de ver, aqueles en que *advertir* era o núcleo verbal, pois ao uso de forma dativa por acusativa con complemento directo cando o referente é masculino tiñamos que xuntar a mestura das dúas estruturas sintácticas coas que é habitual construílo:

a) [Alguén] advirte [algo] [a alguén], onde o pronome desempeña a función de complemento indirecto e, polo tanto, aparece en dativo. Un exemplo:

"Ou seu partido administra mal o resultado, non é dialogante, non é capaz de pactar, convertendo en inútil o voto de moita xente", advertiulle o portavoz do BNG ao líder da oposición. [CORGAetq: DLV2005-05-15/17] Rad3ms

b) [Alguén] advirte [(a) alguén] [de/sobre algo], onde o pronome desempeña a función de complemento directo e, en consecuencia, aparece en acusativo:

Xa a iso das catro o camareiro advertiuno de que eran horas de pecha-lo bar e de que sería millor que se fose deitar. [CORGA: A. Conde, *A casa de Adara*, Xerais, 1996]

Pois ben, consideramos que estes *lle(s)* que proliferan sobre todo nos últimos anos na prensa e que documentamos no corpus de adestramento xornalístico –véxanse os exemplos seguintes– son produto do que en castelán se denomina *leísmo* e mais da mestura das dúas construcións anteriores, o que provocou tamén que os etiquetamos como acusativos:

"Eu entendo que o señor Quintana diga o que ten que dicir", respondeu Núñez Feijoo, antes de advertirlle de que as súas acusacións de voto inútil eran "unha falta de respecto aos galegos; [CORGAetq: DLV2005-05-15/17] [lle Raa3ms lle]

E esas deputadas... ista, ista, ista, ¿non habería que advertirlles da falacia? [CORGAetq: GH2007-03-17/81] [lles Raa3fp lle]

Precisamente, recoñeceu que Martínez Robles co que asegurou que mantén unha "gran relación" xa lle advertira da imposibilidade do proxecto nos "contactos regulares" que mantiñan. [CORGAetq: GH2007-03-17/13] [lle Raa3ms]

O presidente da Xunta, Emilio Pérez Touriño, advertiulle onte ó Ministerio de Fomento de que Galicia "non pode esperar máis" polo Tren de Alta Velocidade de conexión de Galicia coa

Meseta, e recordoulle que está comprometido cuns prazos que deben ser cumpridos.
[CORGAetq: GH2007-03-30/7] [lle Rad3ms lle]

O texto aprobado so fala de traballar pola "desaparición progresiva de símbolos" nos edificios públicos (tamén os crucifixos e a Biblia nas tomas de posesión), sen provocar divisións, e advírtelle á Igrexa Católica da caducidade dos seus privilexios. [CORGAetq: ANT2008-07-10/74] [lle Raa3fs lle]

Cando retomamos o comportamento dos clínicos de terceira¹⁶⁰ para incluílo no protocolo lingüístico e afrontamos o seu estudo, non caso a caso como nos obriga a desambiguación manual, senón no seu conxunto, démonos conta, non obstante, da inconsistencia da nosa actuación.

Por unha banda, é verdade que segundo a normativa oficial e mais o noso coñecemento como falantes o pronome que debería aparecer é o de acusativo: *levouno*, *advertiuno/na*, *instándoo*, etc; ou ben nos casos de *advertir* proceder a eliminar a preposición *de* que precede ao complemento de réxime:

"Eu entendo que o señor Quintana diga o que ten que dicir", respondeu Núñez Feijoo, antes de advertirlle que as súas acusacións de voto inútil eran "unha falta de respecto aos galegos;

o axente preguntoulle se posuía carné de conducir, e ó responderlle afirmativamente, advertiulle que ía perder tres puntos.

...

Mais, por outra banda, debemos ter presente que en verbos coma *axudar*, *obedecer*, *servir*, etc. a alternancia entre as formas pronominais derivadas do dativo latino e as procedentes do acusativo xa se constata no galego medieval (López Martínez, 1993: 189-202), á súa vez testemuño da dobre posibilidade de réxime que posuían no latín.

Nós quixemos comprobar se esa vacilación se seguía dando no galego moderno e decidimos facer unha cala co verbo *informar*. Así mesmo, pareceunos interesante tamén comprobar se casos tipo "instarlle a" estaban xeneralizados xa en galego ou se eran

¹⁶⁰ Apoiándonos na diferenciación do galego entre *o* e *lle*, nos casos en que non hai posibilidades de diferenciar na expresión entre acusativo e dativo (*me*, *nos*, *vos*) optamos por un, a non ser que haxa algún elemento no contexto que nos impida esa caracterización. Así, a conmutación polo dativo de terceira no segundo exemplo está apoiada pola presenza do complemento directo en forma de subordinada substantiva.

Se me buscan, aquí estou (se non me atrevo a desconectar ou quedar fóra de cobertura).
[CORGAetq: ANT2005-03-03/25] [me Raa las me]

Outra, porque me permite abaratar custos na miña empresa. [CORGAetq: GI1996-06/31] [me Rad las me]

produto da tradución automática. Con este fin consultamos o CORGA etiquetado e mais o CORGA, confrontando ademais a prensa coa ficción para ver

- 1) se a alternancia se produce a favor do acusativo ou do dativo;
- 2) se na expresión do complemento predomina a presenza da preposición *a* ou a súa ausencia;
- 3) se as mesmas estruturas se producen tanto na ficción como na prensa;
- 4) se ten base a hipótese da tradución automática, non revisada posteriormente, opoñendo na prensa aqueles xornais que se presume que se traducen (*Galicia Hoxe*), fronte aos medios que se presume que non o fan (*A Nosa Terra* ou *Tempos Novos*).

Á luz da casuística completa das estruturas “informar (a) alguén de/sobre algo” e “instar (a) alguén a/para algo”, extraída do noso corpus e clasificada segundo os parámetros arriba mencionados, incluída no Anexo F deste estudo, podemos destacar que:

A) INFORMAR (A) ALGUÉN DE/SOBRE ALGO

1) No corpus de adestramento de ficción non se localiza ningún exemplo con dativo, fronte ao xornalístico no que atopamos un:

Aínda que as reivindicacións duns e doutros son diferentes a suma de ambas as protestas afectou a numerosos pacientes, que onte se queixaban fundamentalmente de que ninguén lles informou da xornada de paro. [CORGAetq: GH2007-03-30/8]

Polo que se refire ao CORGA, a balanza inclínase na prensa claramente a favor da pronominalización en forma acusativa (35 fronte a 9), e na ficción faino tamén case de xeito exclusivo (98 fronte a 4).

Entre os exemplos de dativo destacan dúas combinacións nas que non é posible a aparición da forma acusativa:

i) Nas impersoais con *se*:

Na orde en que lle comunicaron a custodia do preso, que faría na compañía do cabo Herbal, de volta ao seu destino en Galicia, informábaselle de que se trataba dun "destacado elemento desafecto ao réxime", condenado a cadea perpetua. [CORGA: M. Rivas, *O lapis do carpinteiro*, Xerais, 1998]

Os tres membros do Consello de Administración, que rectificaron partes das súas anteriores declaracións ante a xuíza Teresa Palacios ante preguntas das acusacións, dixerón que se lles

informara dos contratos antes do primeiro Consello de Administración do banco fusionado e que os viron entre os documentos que se acompañan a estas reunións. [CORGA: GH2005-02-08/50]

"O venres ás 11.30 xa se lle informara a todos os portavoces e ó presidente da Comisión de que se retirara porque non tiña sentido volver a formulala". [CORGA: GH2007-03-20/15]

ii) Naqueles casos en que está explícito o complemento e este se refire a organismos ou responsables de organismos, onde se percibe que o trazo + animado está difuminado:

O concelleiro de Seguridade Cidadá de Villajoyosa, Pedro Alemany, informoulle a Efe de que a citada residencia só é ocupada por persoal da entidade bancaria e familiares entre os meses de maio a outubro. [CORGA: GH2005-02-28/29]

2) Cando non existe pronominalización pero si explicitación do complemento, este aparece maioritariamente sen preposición.

No corpus de adestramento xornalístico só documentamos 2 casos sen preposición e 1 con ela, fronte aos textos seleccionados para o adestramento da ficción onde non se recolle ningún.

A prensa do CORGA amosa a alternancia das dúas construcións con valores bastante próximos (12 sen preposición e 8 con ela), aínda que quizais poderíamos aventurar, fixándonos no aspecto cronolóxico, a tendencia cara ao uso con preposición nos últimos anos. Tal orientación non se confirma cos datos que nos achega a ficción do noso corpus, pois só recollemos un exemplo no que aparece explicitado o complemento e, malia ser con preposición, este remítenos á década dos setenta:

Entón non te demores, vaite, que eu informaréi á túa nai. [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

B) INSTAR (A) ALGUÉN A/PARA ALGO

1) No corpus de adestramento de ficción non se rexistra o verbo *instar*, mentres que no xornalístico non se documenta ningún caso con pronominalización en forma de acusativo e si dúas con dativo, separadas cronoloxicamente por unha década (1997, 2007). Así mesmo, a alternancia entre complemento sen preposición e con ela é parella (4 vs. 3).

Nas revistas incluídas no CORGA tampouco aparece ningún exemplo de pronominalización en acusativo e o de dativo é o que rexistramos no corpus de adestramento xornalístico:

A cantidade obxectada deberase ingresar nalgún proxecto alternativo (non se nega a capacidade recadatoria do Estado, só a recadación para fins militares), e o comprobante deste ingreso inclúese na documentación da declaración do IRPF xunto cunha carta dirixida ó Delegado da Facenda expresándolle os termos da obxección fiscal, e instándolle a que no futuro os impresos inclúan a liña da obxección fiscal para recoñecer e facilitar este dereito. [CORGAetq: *TN1997-06/7*]

Agora ben, as revistas amosan un dato que pode ser concluínte: dende o ano 1996 a 2008, período temporal que abranguen os exemplos atopados, en todos os casos en que aparece explicitado o complemento, este constrúese coa preposición *a*.

Curiosamente, a ficción do CORGA escatima o emprego deste verbo, xa que as súas ocorrencias non chegan á ducia. Ademais, aparece utilizado non só na construción transitiva con complemento de réxime que estamos estudando, senón tamén nunha construción que non recollen polo de agora as obras lexicográficas galegas actuais mais que si aparece rexistrada tanto nas do portugués como nas do castelán¹⁶¹. Trátase de “instar algo a alguén”, con réxime idéntico ao de *pedir*:

_Veña esas novidades instoulle o xornalista. [CORGA: A. C. Malvar, *Unha noite con Carla*, Xerais, 1995]

Rexedor: Instarémo-la restitución deses bens e a inmediata liberdade do voso irmán leigo. [CORGA: A. Magán, *Os rebertes. Retablo escénico en catro cadros*, Xerais, 1993]

Este novo réxime tamén se recolle máis recentemente nun exemplo da prensa

O mesmo artigo dispón que cando un partido sospeitoso de ser sucesor doutro ilegalizado solicite a súa inscrición "poderán instar o pronunciamiento da sala sentenciadora o Ministerio do Interior e o Ministerio Fiscal". [CORGAetq: *GH2007-03-30/52*]

o que nos indica que no galego tamén amplía o seu significado.

Volvendo á estrutura que nos interesa aquí, nos textos ficticios do CORGA só se recollen a pronominalización en forma acusativa e mais o complemento explícito coa preposición *a*.

A prensa é bastante máis prolífica na utilización deste verbo e tamén bastante máis variada na súa expresión. Así, nos casos de pronominalización predominan as formas de acusativo (20) fronte ás de dativo (4), nas que en dúas, como xa sucedía con

¹⁶¹ No *Novo Aurélio Século XXI* (1999) *V.t.d.* 1. Pedir, solicitar, com instância: *A carta instava que atendese o pedido da visita*. En María Moliner (2007). 1. tr. e intr. *Insistir en una petición: 'He instado la pronta solución de mi expediente'.

informar, a presenza do *se* marca de impersonalidade e a posposición do complemento + animado difuminado condicionan o emprego do dativo:

Por este motivo, catro deputados do Grupo Parlamentario Socialista (Manuel Vázquez, Ceferino Díaz, José Carlos Baños, Ismael Rego) e o seu voceiro, presentaron unha proposición non de lei instándolle á Cámara galega a esixir do Ministerio de Obras Públicas un compromiso público a este respecto. [CORGA: CG1995-02-10/19]

A Mesa Nacional de HB decidiu remiti-las dúas copias do vídeo de ETA que estaban no seu poder, tras recibir o venres, á unha e media da tarde, a notificación do xuíz Baltasar Garzón na que se lles instaba a entrega-los exemplares do citado documento audiovisual que posuían. [CORGA: CG1996-02-18/55]

Polo que respecta á explicitación do complemento sen estar acompañado de pronominalización, a alternancia entre o uso da preposición *a* e a súa ausencia inclínase cara a esta última, posto que as ocorrencias sen preposición case dobran os casos en que aparece: 206 fronte a 104. Non obstante, malia individualizar os xornais e fixármolos no ano de publicación, non podemos extraer ningunha outra conclusión pois os datos, como se pode observar pola táboa que reproducimos a continuación, non nolo permiten:

<i>Distribución dos casos de “instar a” nos xornais do CORGA</i>			Complemento sen preposición	Complemento con preposición
Xornal	Ano	Palabras por xornal/ano	Número de casos: 206	Número de casos: 104
CG	1995	878 992	24	5
VG	1995	140 693	1	8
CG	1996	929 129	23	7
VG	1997	21 939	0	4
CG	1999	632 006	24	6
CG	2000	1 609 671	67	13
CG	2001	476 527	10	7
ANT	2002	368 899	1	6
CG	2002	643 111	19	8
ANT	2003	222 648	0	2
GH	2005	627 149	10	19
DLV	2007	69 976	1	3
GH	2007	637 974	24	14
GH	2008	90 583	2	2

A modo de resumo, recolleemos na táboa que segue os datos numéricos relativos ás dúas construcións descritas dos verbos *informar* e *instar* na ficción e prensa do CORGA:

<i>Fonte</i>	<i>Verbo</i>	<i>Pronome en acusativo</i>	<i>Pronome en dativo</i>	<i>Complemento sen preposición</i>	<i>Complemento con preposición</i>
adestramento xornal	informar	4	1	2	1
	instar	0	2	4	3
adestramento ficción	informar	3	0	0	0
	instar	0	0	0	0
revistas CORGA	informar	4	2	0	0
	instar	0	1	0	6
xornais CORGA	informar	36	9	12	8
	instar	19	4	206	104
ficción CORGA	informar	98	4	0	1
	instar	4	0	0	4

Os datos que nos achega a cala realizada e mais a alternancia xa coñecida entre as formas de acusativo e as de dativo con verbos coma *informar*, *axudar*, *obligar*, etc. lévannos á conclusión seguinte: hai verbos que, historicamente na nosa lingua mais tamén noutras coma castelán ou portugués, amosan unha vacilación no uso dos clíticos. Este feito, a alternancia observada entre a aparición do dativo e a de acusativo, xunto coa posible contaminación nalgúns casos do castelán, esixiría unha análise de cada caso e mais a explicación pertinente. Se isto o trasladamos á análise automática, incrementaríase a ambigüidade e o problema sería moito máis complexo. Ademais, e sobre todo, o sistema de recuperación de datos sería pouco transparente. Así pois, malia que nalgunhas desas ocorrencias sexa inadmisíbel a aparición do dativo, debemos caracterizar o caso do pronome en correspondencia coa súa expresión:

forma acusativa → etiqueta de acusativo

forma dativa → etiqueta de dativo

Isto obríganos a corrixir as trece análises nas que unha forma de dativo de terceira foi caracterizada na versión 2.4 do *CORGAetq* como acusativa¹⁶², mais a cambio gañamos en transparencia, facilitamos a recuperación por forma e non aumentamos a ambigüidade potencial.

4.3.2.6. Vacilacións ~ variacións na adscrición categorial

Hai unha serie de unidades para as, que segundo a norma vixente e as obras lexicográficas, só é posible unha categorización. Por exemplo *cara* debe remitirse á clase substantivo

"O venres á noite estiven ceando con Blanco e Touriño, pero ningún dos dous me dixo nada", comenta sen quitar o sorriso da cara. [CORGAetq: ANT2008-07-10/13]

¹⁶² Son as seguintes:

1. Recordou que desde a antiga oficina principal, Caixa Galicia iniciou unha expansión pola cidade olívica que lle levou a contar na actualidade cun total de 25 oficinas que proximamente se verán aumentadas en número. [CG1996-02-15/38]
2. A cantidade obxectada deberase ingresar nalgún proxecto alternativo (non se nega a capacidade recadatoria do Estado, só a recadación para fins militares), e o comprobante deste ingreso inclúese na documentación da declaración do IRPF xunto cunha carta dirixida ó Delegado da Facenda expresándolle os termos da obxección fiscal, e instándolle a que no futuro os impresos inclúan a liña da obxección fiscal para recoñecer e facilitar este dereito. [TN1997-06/7]
3. "Eu entendo que o señor Quintana diga o que ten que dicir", respondeu Núñez Feijoo, antes de advertirlle de que as súas acusacións de voto inútil eran "unha falta de respecto aos galegos; [DLV2007-05-15/17]
4. A artista quixo ser nai logo de romper a súa relación co ciclista Lance Armstrong pero un cancro, do que xa está recuperada, obrigoulle a pospoñer, ata agora, o seu desexo. [DLV2007-05-15/36]
5. Precisamente, recoñeceu que Martínez Robles co que asegurou que mantén unha "gran relación" xa lle advertira da imposibilidade do proxecto nos "contactos regulares" que mantiñan. [GH2007-03-17/13]
6. o axente preguntoulle se posuía carné de conducir, e ó responderlle afirmativamente, advertiulle de que ía perder tres puntos. [GH2007-03-17/40]
7. E esas deputadas... ista, ista, ista, ¿non habería que advertirlles da falacia? [GH2007-03-17/81]
8. Aínda que as reivindicacións duns e doutros son diferentes a suma de ambas as protestas afectou a numerosos pacientes, que onte se queixaban fundamentalmente de que ninguén lles informou da xornada de paro. [GH2007-03-30/8]
9. A Estados Unidos séguelle na lista dos países menos amigos Marrocos (5,8 por cento), mentres que tras Francia, entre os mellores amigos, sitúase Portugal (5,6%). [GH2007-03-30/29]
10. O xogador serbio saltou ó Estadio de Cayo Biscaíno moi concentrado en non repetir os erros que lle levaron a perder en Indian Wells e na segunda ocasión de que dispuxo logrou romper o servizo do xogador español para colocarse cun 3-1 ó seu favor. [GH2007-03-30/75]
11. Nadal intentou reaccionar, pero o gran xogo do seu rival levoule a ceder a primeira manga por 6-3. [GH2007-03-30/75]
12. O texto aprobado so fala de traballar pola "desaparición progresiva de símbolos" nos edificios públicos (tamén os crucifixos e a Biblia nas tomas de posesión), sen provocar divisións, e advirtelle á Igrexa Católica da caducidade dos seus privilexios. [ANT2008-07-10/74]
13. "O avogado da actriz rexeitou repetidamente os intentos dos (axentes) federais de interrogala", di o diario, que asegura que, segundo as súas fontes, as autoridades poderían solicitarlle agora "a un gran xurado unha citación para obrigarlle a que lles diga todo o que sabe" sobre o suceso. [GH2008-08-05/48]

cara Scfs cara

ou á clase adxectivo

Galicia e Estremadura teñen a Súper máis cara [CORGAetq: CG2000-11-23/63]

cara A0fs caro

Non obstante, hai ocasións en que calquera desas dúas posibles asignacións sería incorrecta, pois a súa función na secuencia dinos que nin se pode integrar na categoría dos substantivos nin na dos adxectivos. Vexamos unha pequena mostra destes outros casos:

É certo que a economía norportuguesa presenta niveis salariais máis reducidos que a galega (por certo que esta circunstancia non parece ter animado ao empresariado catalán ou navarro para deslocarse atropeladamente cara Galiza) pero non é menos certo que a economía galega, o traballador medio galego, manifesta unha produtividade superior. [CORGAetq: G11996-06/34]

Ollando cara diante, só vemos que o problema se vai agravar. [CORGAetq: ANT2002-10-24/31]

Así mesmo, Adegas manifestou a súa oposición ós dous proxectos e denunciou un suposto "trato de favor" cara Stolt Sea Farm. [CORGAetq: GH2007-03-30/3]

En canto a loxística, ten unha boa situación para ser unha porta cara os mercados internacionais. [CORGAetq: ANT2008-07-10/70]

É obvio que nos exemplos anteriores *cara* exerce como preposición, mais nos manuais non se inclúe na listaxe de preposicións, fronte a *cara a* que si se conta entre as locucións prepositivas. Tamén é obvio que a expresión que debера figurar nas secuencias anteriores é *cara a*; non obstante, posto que non aparece, nós na etiquetaxe manual non reconstruímos a locución, senón que caracterizamos o *token cara* como unha preposición máis, e lematizámolo na forma que presenta no texto:

cara P cara

Naturalmente, incluímos tamén no lexicón esta última opción clasificatoria da unidade *cara* para que o etiquetador automático a poida recoñecer e caracterizar ben cando se etiqueten sen supervisión manual outros documentos do CORGA.

Sen causar ningunha repercusión no sistema a maiores da que xa vimos, o emprego da preposición *cara* cando vai seguida dun adverbio que comeza por *a*, coma en

E hai que remontarse cara atrás ata o ano 1989 para dicir que no nadal que acaba de pasar a xente gastou un pouquiño máis que hai seis anos, pero menos ca en calquera dos últimos trimestres de 1994, 1993, 1992, 1991 e 1990. [CORGAetq: CG1996-02-16/1]

Este é un período moi semellante en fecundidade e vizo a outros intres de salto cara adiante do noso pensamento, como o Rexurdimento, na I República, e a Xeración Nós e o Partido Galeguista, na II República. [CORGAetq: GH2008-07-10/27]

provoca bastante máis controversia entre os estudosos da filoloxía galega. Tanto é así que as opinións están divididas e hai quen, coma Álvarez e Xove (2002: 634), afirma que a preposición *a* desta locución se omite ante os adverbios de lugar deícticos (*aquí, alí...*) ou ante os que xa teñen aglutinada a preposición (*arriba, atrás...*), mentres que outros, coma Hermida (2004: 191), refutan esta tese fundamentándose no uso oral en cuxa transcripción fonética rexistran a aparición da locución.

Algo semellante pero menos usual sucede con *fronte*, forma que tamén clasificamos como preposición en casos coma:

Esta cifra supón unha media de 283 euros por habitante, fronte os 392 do conxunto do Estado. [CORGAetq: ANT2002-10-03/31]

E así foi como un se viu obrigado a aprender a historia de Irlanda, católica e independentista fronte a protestante e asoballadora Inglaterra. [CORGAetq: GH2007-03-30/88]

A deputada popular tamén aludiu ao tipo de axudas e criticou que en Galicia haxa "cero" prestacións recoñecidas por teleasistencia, axuda a domicilio, centros de día e atención residencial fronte outras autonomías. [CORGAetq: GH2008-08-05/38]

4.4. Peculiaridades na etiquetaxe e lematización manuais de...

4.4.1. O pronomes persoal tónico

As formas pronominais tónicas clasifícanse no noso sistema con respecto ao caso en nominativas e preposicionais. En xeral, se as precede unha preposición ou locución preposicional son caracterizadas na etiquetaxe como caso *preposicional*, mentres que se non as antecede ningunha das unidades anteriores considéranse ocorrencias do caso *nominativo*. Por exemplo:

Nominativos:

"Eu estou bastante afeito a escoita-los portavoces do PP no Parlamento loa-lo modo no que traballan e cumpren as súas funcións estes dous organismos", engadiu. [CORGAetq: *CG1996-02-14/89*]

E por enriba de todo elas, nós, tamén temos dereito a participar nas decisións. [CORGAetq: *ANT2005-03-03/67*]

A política é unha arte totalmente nobre para a que eu, sen lugar a dúbidas, non serviría. [CORGAetq: *DLV2007-05-15/20*]

Esta iniciativa de Joseph Blatter (na foto) non ten amparo legal e el mesmo a defendera a través dun "pacto ético". [CORGAetq: *ANT2008-07-10/88*]

Preposicionais:

As cousas invertíronse e é o Madrid quen depende del para gañar a Liga. [CORGAetq: *DLV2007-05-15/111*]

Entrementres, neste cruzamento de opinións que aínda non acabou, ao meu redor máis espectadores comparten comigo unha nova necesidade que se incorpora á lista de demandas vitais: [CORGAetq: *GH2007-03-30/85*]

Trátase de que o mítico animal non se extinga e de actuar contra el só naquelas zonas onde cause danos reiterados. [CORGAetq: *ANT2008-07-10/59*]

E tamén hai solucións se non viaxan connosco. [CORGAetq: *GH2008-08-05/39*]

Naturalmente, se o pronome non é o termo dunha frase preposicional, aínda que sexa a palabra que dun xeito casual segue á preposición, non se caracteriza no caso preposicional senón no nominativo, coma en:

En Puerto Rico, o inglés _que si é oficial xunto co español_ era a lingua na que me falaban antes de eu contestar en castelán. [CORGAetq: *GH2007-03-17/84*]

O dilema preséntasenos nos casos de sincretismo en dobre dirección. Por unha banda, cando o pronome tónico é o segundo termo dunha estrutura comparativa, pois malia que se nos di (Álvarez, Regueira e Monteagudo, 1986: 171; Álvarez e Xove, 2002: 549) que as conxuncións *coma* e *ca* esixen o emprego da forma oblicua, só cando o referente é unha primeira persoa, podemos constatar na expresión esa diferenciación.

Exemplos:

Supoño que pensarán coma min, que o tema ten importancia, porque hai cartos de por medio. [CORGAetq: *GH2007-03-17/17*]

Alberto Fortes recupera en Tristemil unha historia gótica sobre a imaxe do dobre, ese outro que é coma nós sen ser nós. [CORGAetq: ANT2005-03-03/40]

Por outra banda, non todas as preposicións esixen que o seu termo sexa un pronome tónico oblicuo, posto que as que podemos denominar exclusivas (*aga, agás, bardante, excepto, fóra, menos, quitado, quitando, sacado, sacando, salvo*) rexen formas rectas:

"Ninguén, agás eu, pode saber xa como caía a chuvia aquela noite na estación de B.". [CORGA: D. Pérez Iglesias, *Estación Término*, Xerais, 1993]

Había dous homes á barra, dous "pailáns" (este termo ofensivo axiña o aprendían os madrileños) aos que ninguén daba atención salvo eu, que os coñecía por seren da aldea: o Piloto, que andaba embarcado, e o Zóupalle-zóupalle, paxeiro. [CORGA: X. Alcalá, *A nosa cinza*, Follas Novas, 1980]

Tampouco sei se alguén máis, quitado eu, pasou por este calvario. [CORGA: X. Borrazás, *Na maleta*, Sotelo Blanco, 2000]

Nesta excepcionalidade tamén se inclúen as consideradas preposicións modais, (*segundo, conforme e consonte*), mais no CORGA non atopamos ningún exemplo de ningunha delas seguida única e exclusivamente polo pronome de primeira singular, ou sexa, os usos que encontramos son os que se corresponden coa conxunción introdutora dunha subordinada modal:

O verdadeiramente importante, segundo eu o vexo, é que os galegos, ou os de calquera outra parte, sexamos libres, teñámo-la oportunidade de escolle-la nosa propia vida e poidamos vivila con dignidade. [CORGA: M. Vázquez Freire, *Anxos en tempos de chuvia*, Xerais, 1997]

O garda sentado ante a mesa pequena tecteaba na máquina de escribir conforme eu ía respondendo. [CORGA: C. Gonsar, *Desfeita. (Semirreportaxe)*, Xerais, 1983]

Si se documentan, non obstante, as preposicións modais con pronomes pertencentes a outras persoas, mais estes usos non nos valen porque neles batemos co problema do sincretismo formal:

Caruana explicou que tralos atentados en Estados Unidos aumentaron as incertezas e, consonte el, se ben é difícil medi-la magnitude da desaceleración económica mundial, o que parece máis claro é que existen "riscos de revisións á baixa" dos indicadores de crecemento. [CORGAetq: CG2001-10-09/32]

Segundo ti _continúa o xefe interesado_, debemos pensar en persoas relacionadas coa química, coa medicina... [CORGA: M^a. X. Queizán, *Ten o seu punto a fresca rosa*, Xerais, 2000]

No TILG tampouco documentamos ningunha ocorrencia de *segundo* / *conforme* / *consonte* como preposicións cuxo termo sexa o pronome *eu*¹⁶³.

Outra posibilidade, xa que semella misión imposible atopar un caso co pronome de primeira nos corpus do galego¹⁶⁴, sería, xa que logo, averiguar qué forma empregan os falantes que teñen *tu* como nominativo e *ti* como preposicional, mais nos corpus consultados (CORGA, TILG e CTAG), malia haber textos ou transcricións dialectais con esas características (CORGA e TILG), non se dá a combinación que buscamos.

É verdade que podemos botar man das dúas linguas máis próximas á nosa para comparar os resultados, pois en ambas as dúas, portugués e castelán, tanto o pronome de primeira como o de segunda singulares posúen formas distintas segundo funcionen como nominativo ou como preposicional: *eu, tu* / *mim, ti*; *yo, tú* / *mí, ti*. Así, na consulta ao *Corpus del Español* de Mark Davies¹⁶⁵ atopamos entre outras ocorrencias:

Por supuesto, enmudecí, aunque quería hacer preguntas. No. Quería respuestas. Y no es que no lo creyera; no entendía. Según yo, un infarto era sinónimo de muerte rotunda. [Corpus del español: C. Miranda, *Los Cenicientos o el Infarto de mi Mamá*]

Eso lo sabemos bien los pobres. Pero, según tú, parece que Jesucristo lo que quería era una sociedad de pobres, de indigentes y de mendigos. [Corpus del Español: P. Chamizo, *Paredes, un campesino extremeño*]

Tamén no *Corpus de referencia del español actual* (CREA)¹⁶⁶ se rexistran casos de ambas as persoas en nominativo:

¹⁶³ Desafortunadamente, tampouco o CTAG nos serve de moita axuda, xa que non rexistra ningún caso de *conforme* ou *consonte* como preposición. Así mesmo, para *segundo* non discrimina na práctica entre preposición e conxunción, xa que, segundo os resultados obtidos nunha consulta realizada o 03/05/2011 [http://sli.uvigo.es/CTAG/ctag_pescuda.php?pescuda=segundo&tipo_busca=palabra&categoria=CONX], non se obtivo ningún resultado sobre a pescuda da palabra *segundo* coa categoría gramatical de conxunción e si 1490 coa mesma palabra coa categoría preposición, dos que o primeiro é:

Este incremento puede semellar pequeño, mais os seus efectos poderían ser realmente temibeis, segundo nos di maioría da comunidade científica.

¹⁶⁴ Probablemente porque colocalo como termo da frase preposicional pode interpretarse como sinal de prepotencia ou descortesía. No apoio desta hipótese está o uso do plural de modestia ou perífrases como “na miña opinión”, “ao meu parecer”, “para min”, etc. Sería posible tal uso, non obstante, en enunciados moi insistentes ou repetitivos nos que se pretende destacar o contraste entre dúas opinións.

¹⁶⁵ Disponível en <http://www.corpusdelespanol.org> [consulta realizada o 05/05/2011].

¹⁶⁶ Accesible en <http://www.rae.es/creanet.html> [consultado o 05/05/2011].

Y es que no te hartas, ¿eh? Sales una mañana y vuelves de la playa con un medio ahogado, según tú. [CREA: A. Gala, *¿Por qué corres, Ulises?*, Espasa Calpe, 1994]

pero mi indignación y mi estado de ánimo personales no tenían por qué verse burdamente al desnudo en mi cuento, el cual, según yo, debía ser frío, carente de emoción y casi indiferente, para su mayor eficacia. [CREA: A. Monterroso, *Literatura y vida*, Alfaguara, 2004]

Polo que respecta ao portugués, no *Corpus do Português* de Davies e Ferreira¹⁶⁷ demos facilmente, por exemplo, con:

Foi aí que entrou D. ou o assassino, segundo L, segundo a sobrinha da anfitriã, ou A, segundo eu mesmo, que ouvia tudo boquiaberto. [Corpus do Português: B. Carvalho, *Os Bêbados e Os Sonâmbulos*, 1996]

Segundo tu, devíamos estar sempre a pedir desculpa da nossa insignificância. [Corpus do Português: J. G. Simões, *Uma História de Província: Vida Conjugal*, 1936]

Mais cando quixemos ampliar as secuencias exemplificativas en portugués e contrastar o *Corpus do Português* con outras fontes, comprobamos que as dificultades aumentaban. Recorremos primeiramente á busca no CETEMPúblico¹⁶⁸, pero alí non se rexistra ningún caso da preposición *segundo* seguida do pronome recto de primeira ou segunda singular, *eu* ou *tu*, posiblemente polo tipo dos documentos que contén, só xornalísticos. Non obstante, cal non é a nosa sorpresa cando ao consultar o *Corpus de Referência do Português Contemporâneo* (CRPC)¹⁶⁹, un corpus constituído por máis de 330 millóns de palabras pertencentes ao discurso escrito (literario, xornalístico, técnico, didáctico, económico, xurídico, etc.) e discurso oral (alocucións informais e formais), tampouco atopamos ningún exemplo. Temos que acudir ao Projecto AC/DC (*Acesso a corpos/Disponibilização de corpos*), que á parte do CETEMPúblico inclúe outros corpus do portugués, para reafirmar na lingua veciña a existencia de preposición modal cuxo termo é un pronome en forma recta¹⁷⁰:

¹⁶⁷ Disponível em <http://www.corpusdoportugues.org> [consulta realizada o 05/05/2011].

¹⁶⁸ Corpus de Extractos de Textos Electrónicos MCT/Público, disponível em <http://www.linguateca.pt/CETEMPUBLICO/>

¹⁶⁹ A versão 2.0 está disponível em <http://alfclul.clul.ul.pt/CQPweb/crpc/index.php> [consultado o 05/05/2011].

¹⁷⁰ Das oito ocorrencias en que aparece a secuencia “segundo eu”, só unha delas se corresponde coa estrutura que buscamos; pola contra, non encontramos ningún caso en que o termo sexa “tu”. Tampouco logramos averiguar cal dos corpus é o que contén as secuencias nin cal é a fonte concreta na que se documenta o exemplo. Consulta realizada o 05/05/2011 en <http://www.linguateca.pt>.

É essa instrução moral que caracteriza a obra do imortal Richardson e que, segundo eu, dá-lhe vantagem sobre todos os seus rivais: instrução tão mais necessária hoje para nós, que todas as fontes onde a juventude dela tinha se abeberrado estão hoje secas ou corrompidas.

En conclusión, queda demostrado o uso das formas rectas coas preposicións modais en portugués e castelán, uso que cremos lexitimamente extrapolable ao galego. Mais, xa que isto é así, por que cunha preposición modal ou excluínste se rompe a regra xeral que rexe forma pronominal tónica oblicua tras preposición? A resposta pensamos que pode estar na accidentalidade deste tipo de preposicións.

Cunha e Cintra (1984: 552) clasifican, moi acertadamente na nosa opinión, as preposicións en esenciais e accidentais –outros fano en *propias* e *impropias*– opoñendo as que sempre o foron ás que, pertencendo normalmente a outras clases de palabras, funcionan ás veces como preposicións.

Alarcos (1994) mantén:

Más dudoso es el carácter de ciertos adjetivos, también inmovilizados en masculino singular, como *excepto*, *salvo*, *incluso* y los cuantificadores *más* y *menos*. Si bien adoptan la atonicidad propia de las preposiciones y constituyen con el sustantivo a que acompañan un adyacente circunstancial, no son propiamente preposiciones. [Alarcos, 1994: 216-217]

E, ademais, exclúe do inventario de preposicións *según*:

Se excluye *según*, porque es unidad tónica (y no átona como las otras) y puede aparecer aislada, por ejemplo en respuestas: –¿Qué harías en esos casos? –*Según*. Puede combinarse directamente con verbos. Al aparecer ante sustantivos, como sucede con las unidades relativas dependientes (*donde*, *cuando*, *como*, § 138) puede pensarse que el verbo se ha elidido. [...] Asimismo, los sustantivos personales de primera y segunda persona en singular no adoptan precedidos de *según* los significantes que les son propios tras la preposición: no se dice *según mí* ni *según ti*, sino *según yo* y *según tú* (igual que con los citados relativos: *donde yo*, *cuando tú*, *como él*, aunque hay ejemplos americanos como *Lo mandaron donde mí*, 100.187). [Alarcos, 1994: 217-218]

Na nosa etiquetaxe seguimos a tradición galega e, xa que logo, mantemos o lugar que lles corresponde entre as preposicións ás excluínstes (*aga*, *agás*, *bardante*, *excepto*, *fóra*, *menos*, *quitado*, *quitando*, *sacado*, *sacando* e *salvo*) e ás modais (*segundo*, *conforme* e *consonte*), coa salvidade do emprego de *segundo*¹⁷¹ –etiquetado

¹⁷¹ Hai aínda outro uso adverbial, cremos, mais este só o documentamos coa variante castelanizada. Trátase da forma *según* ou *asegún* que aparece en certas respostas características da fala. Por exemplo:

como adverbio– cando actúa como conector discursivo axudando a organizar o discurso. Así, fronte á etiqueta categorial preposicional das secuencias de a) e á etiqueta conxuntiva das de b) opoñemos a adverbial das de c):

a) Preposición

¿E que é o normal segundo ti, guapa? [CORGA: X. A. Neira Cruz, *As cousas claras*, Xerais, 2001]

Este dato confirma que a economía española continuou a traxectoria de desaceleración iniciada no primeiro trimestre do 2000, segundo o INE. [CORGAetq: *CG2002-05-30/5*]

Indo por partes, ese espazo foi ideado para achegar a figura de Zapatero a ese colectivo que formamos todos e que recibe por nome xente, por iso o presidente do Goberno optou por falar de "ti" a cada participante, optando por un cáldo "Lorenzo" para comunicarse cun moderador/xornalista que, en segundo que caso foi demasiado frío cos asistentes coa repetitiva petición de que falaran ben apegados ó micrófono, iso sí, Milá foi quen de facer algunha réplica, tímida, certo pero útil para darlle ritmo a unha proposta que probou a mesma eficacia que na tele gala. [CORGAetq: *GH2007-03-20/100*]

Telecinco renovou o pasado mes de xullo o seu liderado no ránking de Sofres, polo que leva catro tempadas ou catro anos ou 30 meses, segundo como se contabilice, como a cadea máis vista de España. [CORGA: *GH2008-08-06/115*]

b) Conxunción

O prezo da vivenda subiu no Estado español desde 1980 un 124% en termos reais, segundo revela o último estudio do mercado inmobiliario internacional do semanario The Economist. [CORGAetq: *ANT2002-10-10/36*]

A UE debe buscar mecanismos eficaces para loitar contra a "epidemia" de piratería que provoca unha perda de 100.000 postos de traballo anuais, segundo destacaron onte os secretarios de Estado de Xustiza e para a Unión Europea, Jose María Michavila e Ramón de Miguel, respectivamente. [CORGAetq: *CG2002-05-24/11*]

Farto de me xuntar nas tabernas do porto con vellos mariñeiros carcomecidos polas propias historias que narraban, engrandecendo dubidosas fazañas coa axuda de vasos de xenebra. Daquela era un máis, ou un menos, asegún. [CORGA: M. Villar, *Premios Pedrón de Ouro 1987: Verbas cruzadas con Amaranta*, Edicións do Castro, 1988]

Sarxento: (Amable). ¿Que tal a túa noiva? Tristán: (Confuso). Pois... según. [CORGA: X. Cermeño, *A televisión. Como se fai unha serie de televisión*, Xunta de Galicia, 1993]

¿Son os alcaldes altos cargos? Según. [CORGA: *ANT2002-09-19/1*]

En total, segundo informan as ONG que traballan na rexión, calcúlase que son preto de 4,5 millóns as persoas afectadas polo conflito que viron como se destruían os seus medios de vida tradicionais. [CORGAetq: GH2008-08-05/97]

c) Adverbio

Fichar un novo técnico supón, primeiro, pagarlle unha indemnización ao adestrador actual, e, segundo, pagar un salario extra a un novo equipo técnico. [CORGA: ANT2007-02-08/58]

Inés: Primeiro son señorita, non señora, e segundo, non especulo, simplemente seino. [CORGA: J. L. Prieto Roca, *Fobias*, Embora, 2005]

As combinacións do xogo son as seguintes, sendo cada unha delas vencida pola inmediatamente superior: primeiro, a carta máis alta; segundo, unha parella; terceiro, dobres parellas; cuarto, figuras; quinto, trío; sexto, escaleira; sétimo, color; oitavo, full (parella e trío); noveno, póquer (catro cartas do mesmo valor); décimo, escaleira de color; undécimo, repóquer (cando se xoga con comodín). [CORGA: X. C. Caneiro, *A rosa de Borges*, Sotelo Blanco, 2000]

A cuestión, entón, de como caracterizar o caso do pronome tónico cando se produce sincretismo entre as formas rectas e as oblicuas, resolvémola conmutando polos pronomes de primeira singular, xa que no caso desta persoa non existe tal problema. Así guiándonos por esta, e en función das conmutacións correspondentes, caracterizamos o pronome persoal tónico cando conforma o segundo termo dunha comparación con *coma* ou *ca* como *preposicional*, mentres que como termo dunha preposición excluínte ou modal será remitido ao caso *nominativo*.

Por suposto, no caso de se rexistrar nun dos supostos anteriores unha ocorrencia de pronome tónico de primeira singular¹⁷², ben sexa na súa forma recta ben na oblicua, por moito que normativamente non sexa a correcta, nós etiquetaremos o caso en función do significante do pronome. Así, malia que de momento non se deu ningún caso no corpus de adestramento, no CORGA aparecen, entre outros, os seguintes, os cales serán caracterizados como *nominativos*:

¹⁷² Operación aplicable tamén sempre ao pronome de segunda singular se nos textos se distingue nesa persoa entre a forma de nominativo e a preposicional, como, por exemplo, se explicita nas seguintes secuencias:

Por eso prefirinte a tí e non a un crego, porque tu non estás na obriga de calar as miñas immoralidades e podes contalas, responsablemente, craro é. [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

¡Eu quería pra tí o mellor do mundo, e tu es a miña acabación! [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Manolo tiña os mesmos anos ca eu: os dous tiñamos trinta e catro. [CORGA: M. Gondar Portasany, *Mulleres de mortos. Cara a unha antropoloxía da muller galega*, Xerais, 1991]

Si a tivérades que escoitar tantas veces coma eu, daríavos ganas de meterlle un pano na boca. [CORGA: X. Bernárdez Vilar, *Un home de Vilameán. Anatomía dunha revolución*, Patronato da Cultura Galega, 1976]

_Son un indisciplinado, non proba os papeis que fago, non ensaio e o elenco artístico que me acompaña é tan independente e bohemio coma eu mesmo. [CORGA: CG1996-02-10/179]

As peculiaridades formais dos tónicos non rematan aínda, a pesar do que xa vimos, posto que tamén nos causa dificultades a súa combinación coa preposición *entre*.

Segundo a *Gramática galega* (Álvarez, Regueira e Monteagudo, 1986: 172) “*a preposición entre rexe formas oblicuas cando a frase nominal funciona como complemento de lugar e formas rectas cando indica coparticipación na acción: Pono entre ti e min; Facémolo entre eu e ti*”. Freixeiro (2000: 126, 560) concorda con esta obra, e Hermida (2004: 77) mantén tamén o emprego das dúas posibilidades segundo presenten unha significación ou outra.

Pola súa banda, Álvarez e Xove (2002) apuntan así mesmo a especialización con *entre*:

As formas de nom. empréganse trala prep. *entre* para expresar coparticipación na realización da acción; tamén son posibles en SPrep. con *entre* con outras funcións, mais nese caso recoméndanse as formas compl. *Eu*, e non *min*, é a forma empregada no segundo termo da comparación tras *do que* cando se relaciona co Sux.:

As maletas subímolas *entre eu e Duarte*.

Ti aínda es más inocente *do que eu* [= do que son eu]. [Álvarez e Xove, 2002: 549]

Tras estas últimas afirmacións, dúas son as cuestións que merecen reflexión. Por un lado, a desagregación en dúas estruturas significativas e expresivas distintas en función da combinación de *entre* con pronome tónico: se indica cooperación ou participación rexe forma de nominativo, mentres que se reflicte localización espacial ou temporal rexe forma preposicional, polo menos na norma. Por outro lado, o aserto anterior do emprego de *eu* como segundo termo da comparación tras *do que*.

Se comezamos por este último punto, a utilización da forma recta do pronome tónico no segundo termo da comparación cando este segue a *do que*, dámonos conta da inadecuación ou falta de acerto na exemplificación de formas de nominativo en relación

coas preposicións. É verdade que os autores tratan o tema dos casos e funcións dos pronomes persoais, mais, que ten que ver o nominativo *eu* cunha frase preposicional, á parte da xa mencionada con *entre*? Quizais se deba a unha mala lectura pola nosa parte ou a unha redacción inaxeitada se o propósito era destacar, na expresión da comparación, o emprego das formas de nominativo tras *do que* fronte ás preposicionais tras *ca* e *coma*. Agora ben, tampouco entendemos a escolla do exemplo, pois, fronte ao portugués onde si é de regra o uso de *do que* nas comparativas de superioridade e inferioridade, independentemente do tipo de unidade que constituía o segundo termo da comparación (Cunha e Cintra, 1984: 257; Vilela, 1999: 232-235; Mira Mateus et al., 2003: 732-743), os autores coñecen moi ben a realidade lingüística galega e saben da excepcionalidade da forma pronominal como único elemento do segundo termo da comparación con *do que*. Tanto é así que no *Atlas Lingüístico Galego*, no volume correspondente á morfoloxía non verbal, onde se recolle no mapa 141 (*Máis grande ca min*) a distribución dos marcadores da comparación e mais as formas pronominais que os seguen (García e Santamarina, 1995: 160-161) obtéñense os seguintes resultados para *do que*:

do que min: O31 (Lobios)

do que a min: P15 (Pontevedra)

do que eu: C37 (Carnota) e P25 (O Porriño)

No CORGA, descoñecemos se por influxo do portugués, se porque está implícita no segundo termo da comparación a forma verbal presente no primeiro termo ou se por unha combinación de ambos os factores, documentamos 6 casos nos que a forma de nominativo *eu* é o segundo termo da comparación tras *do que*¹⁷³:

Máis novo do que eu, non sei si por receo ou por respecto, na presenza miña amostrábase esaxeradamente retraído. [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Pero, ó mesmo tempo, imaxinéi ó preso máis desamparado de amigos do que eu, i a súa aboa máis desvalida de influencias que miña nai. [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

¹⁷³ A estes casos deberíamos engadir outros seis nos que o segundo termo está formado só por nomes propios, coma en

Era unha cidade grande, maior do que Vigo, paréceme a min. [CORGA: X. M. Martínez Oca, *As florestas do Mañuema*, Xerais, 1988]

Con certeza que o Xeneral Hoope ou calquera outro dos valerosísimos membros do seu Estado Maior deseñarán mellor do que eu o cadro das operacións rexistradas nos pasados días ata a actualidade. [CORGA: X. R. Pena, *O reverso do espello*, Xerais, 1984]

Como vexo que é vostede un home observador e que, sen dúbida, hai xa tempo que os coñece, ou polo menos bastante máis tempo do que eu, quixera pedirlle, sempre e cando non tiver inconveniente en facilitarma, certa información sobre a familia Hoffmann. [CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]

Non sei se me queda pergameo para rematar este relato, non sei se me foi concedido o tempo necesario, non sei se alguén atopará o meu corpo e as letras escritas pola miña man, máis quentes do que eu. [CORGA: X. Miranda, *Historia dun paraugas azul*, Xerais, 1991]

_Estar estivo nesta casa, pero pasen á cociña a tomar un café, que alí lles poden explicar mellor do que eu. [CORGA: H. Puentes, *O bandido Casanova*, Xerais, 2000]

fronte ás 78 ocorrencias nas que no segundo termo o pronome (57) ou o antropónimo (21) é o suxeito dunha cláusula explícita. Por exemplo:

O tempo é tan tirano que non se pode vivir á súa marxe e agora descubro que corre máis do que eu cavilei cando marchaba canda el. [CORGA: CG2001-10-12/33]

Se pensas que vou facer semellante cousa é que estás máis tola do que eu pensaba. [CORGA: R. Vidal Bolaño, *Saxo tenor (Desgracia arrabaldeira improbable entre loucos, chourizos, gánsters, chulos, currantes e putas)*, Xerais, 1993]

"A chegada á mansión fora moito máis feliz do que eu podía sospeitar... [CORGA: M. Lourenzo, *Nocturnos*, Laiovento, 2005]

Daquela, veciños, se con todo iso non acadastes o perdón, aínda fostes peores ca min, peores do que eu supuña, e supuxen moito; [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Camiñaron horas, e o cayo resultou máis estenso do que Pedro supoñía. [CORGA: X. Neira Vilas, *Galegos no Golfo de México*, Edicións do Castro, 1980]

A xente aplaudiu, máis do que Lamote supuña, algún tempo. [CORGA: P. Martín, *Das cousas de Ramón Lamote*, SM, 1992]

Concluído o mes de xullo, agosto corría moito máis do que Miguel desexaba. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

Era máis do que Constanza podía soportar. [CORGA: M. Gándara, *Magog*, Xerais, 1997]

Curiosamente, *do que eu*, sen aparecer en cláusula, non se utiliza no CORGA en textos posteriores ao ano 2000 nin se rexistra ningún exemplo na prensa; todos os casos se dan na ficción, o que non deixa de ser, cando menos, interesante. Á hora de etiquetar

a forma do pronome non nos afecta a estrutura do segundo termo da comparación, pois atémonos ao significante do pronome, e polo tanto este caracterízase como *nominativo*.

Máis atraente ca o punto anterior é, na nosa opinión, a selección que en teoría efectúa a preposición *entre* co seu termo e a posición do pronome de primeira na coordinación que acostuma seguila.

Na teoría, como vimos máis arriba, dise que a forma recta segue *entre* cando indica cooperación/coparticipación (1a, 1b), mentres que a forma preposicional séguese cando indica localización espacial ou temporal (2a, 2b)¹⁷⁴.

- 1a) Isto queda entre ti e eu, ¿está clariño? [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]
- 1b) Cómo soubo Helena Zumálave da historia entre eu e Raúl en Berlín é un asunto moi simple. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]
- 2a) O buraco entre ti e min aumentaba sen remedio. [CORGA: R. Raña, *O crime da rúa da moeda vella*, Xerais, 1993]
- 2b) Mal que ben, existe unha barreira defensiva, a barra, que se ergue entre min (ou nós, cando non estou só) e os da outra banda, porque de non ser así hai tempo que estaría vencido, dominado. [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

Non obstante, na práctica, se nos fundamentamos nos textos do CORGA, asistimos á confusión entre esas dúas estruturas sintácticas e mesmo semella que avanzamos cara á desaparición da forma de complemento tras *entre*. Obsérvese, senón, nos exemplos extraídos da versión 1.5 do CORGA, a supremacía da forma pronominal de nominativo (35) fronte á de complemento (8):

A) Ocorrencias de nominativo con *eu* como primeiro elemento da coordinación¹⁷⁵:

¹⁷⁴ É posible que esa diferenciación se estea estendendo tamén á preposición *contra*? Na escrita non rexistramos ningunha ocorrencia coa forma de nominativo, mais na fala, polo menos no noso idiolecto, non nos é estraño, co significado de asociación/cooperación, o emprego de *contra eu e ti*, sempre que existan dous bandos manifestos. Por exemplo, *Botamos unha brisca? E como imos? Fulano e Mengano contra eu e ti / contra ti e eu*.

¹⁷⁵ Unha cuestión que nos intrigou abondo, malia non afectar a etiquetaxe que practicamos, é a posición que ocupa a forma recta do pronome de primeira nunha coordinación, ben con outros pronomes ben con nomes propios ou outras unidades máis complexas. Como podemos ver polos exemplos recollidos, a igualdade na distribución coas formas de complemento (3 para *entre min e x* fronte a 4 para *entre x e min*) contrasta coa supremacía observada nas formas rectas da última posición para o pronome de primeira (26 casos de *entre x e eu* contra os 9 de *entre eu e x*). No castelán a norma relega a posición do pronome de primeira nunha coordinación ao último lugar por motivos de cortesía e urbanidade (RAE, 2009: 2408-2409); o *Corpus del Español* de Davies demostra o asentamento desta norma ao rexistrar en todos os casos a construción *entre tú y yo*, malia que no CREA unha presa de casos neguen a rotundidade da

afirmación (exs.: *Y entonces me restregaría así. Y me dirías sí. Y yo y tú. Mi amor. Tus manos aquí. / Era como si yo y él nos hubiésemos ubicado mutuamente por encima de las batallas del amor.*). En portugués, a pesar de que Álvarez (1999: 125) afirma que é indiferente a orde, pois segundo ela “*non rexe o precepto social do español e do francés, que fai coloca-la primeira persoa no derradeiro lugar da estrutura coordinada*”, Cunha e Cintra especifican (2005: 289):

Quando no sujeito composto há um da 1.ª pessoa do singular (eu), é boa norma de civilidade colocá-lo em último lugar:

Carlos, Augusto e eu fomos promovidos.

Se, porém, o que se declara contém algo de desagradável ou importa responsabilidade, por ele devemos iniciar a série:

Eu, Carlos e Augusto fomos os culpados do acidente.

Na realidade, tanto nas consultas ao *Corpus de Referência do Português Contemporâneo* (CRPC) como nas do *Corpus do Português* (CdP), comprobamos que na práctica, non obstante, non se acata esta recomendación, xa que así nolo demostran os resultados:

	CRPC	CdP
<i>entre mim e ti</i>	1	11
<i>entre ti e mim</i>	1	1
<i>eu e tu</i>	27	11
<i>tu e eu</i>	37	21
<i>eu e você</i>	9	7
<i>você e eu</i>	8	9

En galego só Álvarez (1999: 138) se pronunciou sobre esta cuestión, da que considera que “*a norma social de situar no derradeiro lugar a Plsg., imposta desde o ámbito español, é contraria á nosa gramática histórica e á tendencia espontánea da maior parte dos galegofalantes actuais*”. Nós concordamos en que o feito de corrixírennos a colocación do pronome de primeira nunha coordinación coa cantarela aquela de “*o burro diante pra que non se espante*” sendo xa mozotes, cando en teoría a estrutura sintáctica da oración debería estar máis que asentada, contribúe a afirmar a preeminencia das formas de primeira, cuxa xustificación podería orixinarse na orde lóxica da persoa gramatical: 1ª, 2ª e 3ª. Agora ben, como xa demostramos, non podemos apoiarnos nin no portugués nin no castelán para soste esta hipótese e, por riba, dous dos corpus de galego moderno, o TILG e o CORGA, tampouco nos permiten decantarnos pola primeira posición para *eu*, xa que, sexa por influencia do castelán sexa por unha cuestión universal de urbanidade, os datos avalan a norma implícita en manuais de estilo de colocar o pronome tónico en uso recto de primeira singular en último lugar nunha serie. Vexamos que información nos fornecen os corpus:

O TILG (datos facilitados o 16/05/2011) recolle 347 casos co pronome de primeira en última posición (*x e/i eu*), fronte a 286 casos nos que ocupa o primeiro lugar (*eu e/i x*). Desafortunadamente, non podemos conectar estas cifras con datas nin con tipos de texto. Con tipos de texto non é posible porque esa información non se inclúe na codificación; agora ben, a non interconexión con referencias temporais require unha explicación polo miúdo dos motivos polos que non a realizamos. A razón é a seguinte: ante as dificultades amosadas polo sistema de recuperación do TILG localizado na páxina da USC, os datos fóronnos facilitados persoalmente polo profesor Antón Santamarina, labor polo que lle quedamos moi agradecidos; mais a data figuraba dissociada das ocorrencias, de tal xeito que era inviable ir seleccionando nun arquivo os casos correctos por un lado e buscando no corpus ou no outro arquivo a información pertinente a cada caso. Logo soubemos dunha nova versión do TILG na páxina do SLI de Vigo (<http://sli.uvigo.es/TILG/>), pero ao realizar as consultas comprobamos que, por unha banda o contexto amosado non sempre era suficiente para a comprensión da secuencia que buscábamos, e pola outra, tiñamos que ir extraendo cada unha das ocorrencias correctas a outro arquivo para logo traballalas, co que decidimos prescindir dos datos cronolóxicos por ser este un aspecto insignificante na etiquetaxe, aínda que interesante, en relación co traballo que estamos realizando.

O CORGA, quizais polo período temporal que abrangue ou quizais debido á falta dun subcorpus oral, é moito máis concluínte nos seus datos; así, fronte a 146 ocorrencias de *eu* na primeira posición nunha serie, ofrece 1216 nas que esa mesma forma pronominal ocupa o último lugar. Dado que o sistema de

1. Fixo un pequeno escándalo no salón que, entre eu e Agueda, conseguimos acougar e, logo, decidiu que a súa filla marchase de viaxe durante uns días. [CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]

consultas permite realizar as buscas tendo en conta non só períodos temporais senón tamén tipo de documentos, decidimos clasificar os resultados atendendo a ambos os criterios por se se producisen cambios significativos nalgún dos bloques. Así, optamos por separar na ficción os textos dramáticos dos novelísticos e de relato curto, por ser os primeiros os que máis se poden achegar á oralidade; tamén desagregamos na non ficción a prensa do ensaio, por considerarmos nos últimos un estilo máis coidado e elaborado, malia que aquí entran tamén as biografías e entrevistas que, igual ca o teatro, destacan polo estilo directo e o narrador en primeira persoa. Polo que se refire aos períodos cronolóxicos decidímonos só por tres, un por cada década dende o 1975 e englobando ata o 2008 na última delas. Cando se observen os gráficos elaborados cos datos extraídos do CORGA, debe terse en conta que as porcentaxes de palabras por lustros fixadas no seu deseño non son equitativas, senón que se lle concedeu máis representatividade aos últimos períodos, motivo este polo que vemos necesario especificar o volume do subcorpus considerado en cada unha das décadas: 1975-1985, 2 234 462 palabras; 1986-1995, 7 711 127 palabras e 1996-2008, 15 400 953 palabras. Vexamos deseguido os datos que nos proporciona o CORGA clasificados tal e como vimos de describir e mais as gráficas elaboradas a partir deles:

	eu e/i [mais]....				...e/i [mais] eu			
	teatro	ficción	ensaio	prensa	teatro	ficción	ensaio	prensa
1975-1985	1	27	0	0	3	101	1	0
1986-1995	8	21	20	0	65	270	177	12
1996-2008	8	42	2	17	47	440	12	82
Totais	17	90	22	17	115	811	190	95

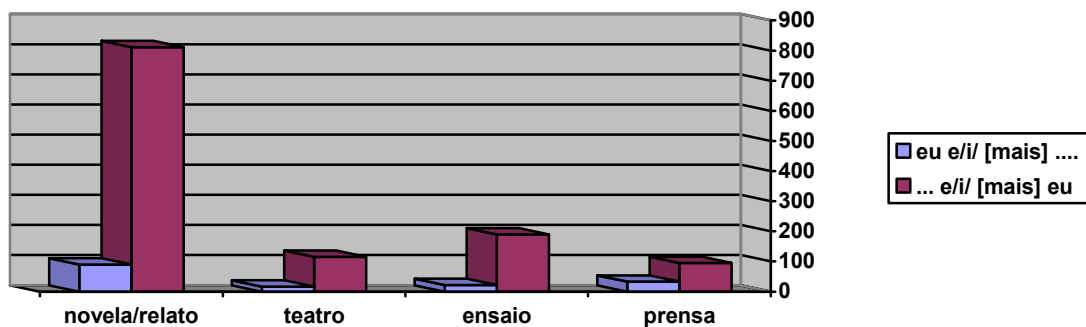


Gráfico 1. Distribución do uso de *eu* nunha serie segundo o tipo de documento. Elaboración propia a partir dos datos extraídos do CORGA, versión 1.5 (<http://corpus.cirp.es/corga>),

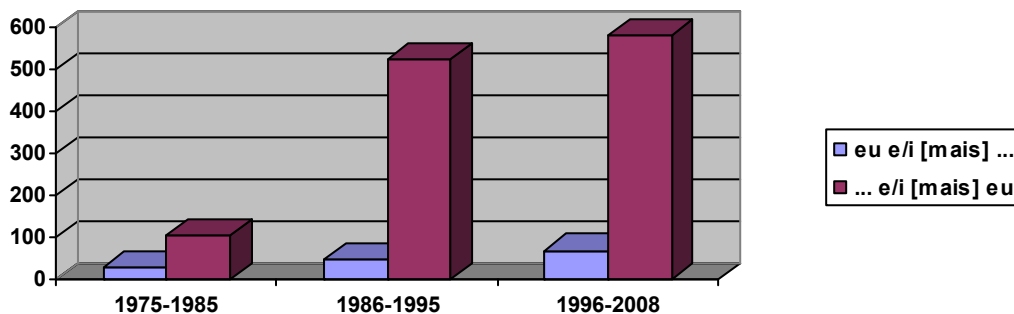


Gráfico 2. Distribución do uso de *eu* nunha serie atendendo ao período cronolóxico. Elaboración propia a partir dos datos extraídos do CORGA, versión 1.5 (<http://corpus.cirp.es/corga>).

2. E todos aqueles trastes ciscados polo chan, entre eu e a porta, entre eu e quen fose. [CORGA: M. Lourenzo González, *Arqueofaxia*, Xerais, 1995]
3. Non había ninguén entre eu e o retrato. [CORGA: X. Alcalá, *Código morse*, Xerais, 1996]
4. Procuo escribir sempre cunha semaniña de adianto, como mínimo, levado sempre da urxencia esta do artigo cotián, do medo a fallar, a faltar un día, levado de calquera ausencia ou de calquera nugalla, mesmo de calquera migraña, ou de calquera outra e asañada circunstancia que se interpoña entre eu e o meu ordenador, entre eu e vostedes mesmos, a través deste espacio do xornal que nunca habería quedar en branco. [CORGA: *CG1999-05-07/10*]
5. Entre eu e aquela máquina había un lazo especial. [CORGA: M. Darriba, *Velada do billarista*, Sotelo Blanco, 2000]
6. Naquel momento, sen embargo, agradecín o traxecto imposto que poñía máis espacio entre eu e os lugares en que adoitaba pasar a práctica totalidade do día, coma se aquel afastamento físico supuxese dalgún xeito ver máis ó lonxe a traxedia que nos tocara vivir da noite para a mañá e que nos tiña a tódolos do instituto e mesmo ós das cafeterías e tendas veciñas que souberan o sucedido co corazón encollido. [CORGA: X. A. Neira Cruz, *As cousas claras*, Xerais, 2001]
7. Pero o meu con el, como diría Ende, xa é outra historia, que quizais non deba ser contada noutro sitio, porque me ensinou a tratar a laranxeira do xardín de atrás e porque ceou comigo unha sobremesa do Renacemento, porque me fixo medrar as pupilas, me deu bicos no antebrazo e paralizou o tempo entre eu e un sofá. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]
8. Cómo soubo Helena Zumálave da historia entre eu e Raúl en Berlín é un asunto moi simple. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]
9. "se o PP pensa que pode desgastarme coa reforma estatutaria terá que demostralo cando o que se decida sexa unha escolla entre eu e eles". [CORGA: *ANT2007-02-22/53*]

B) Ocorrencias de nominativo con *eu* como segundo elemento da coordinación:

1. Resulta, como xa vos dixen, que desde o curso pasado as cousas marchaban mal entre Anxo e eu. [CORGA: M. Suárez Abel, *Premios Pedrón de Ouro 1985/1986*, Ediciós do Castro, 1987]
2. Agora poderá decidi-lo Señor, entre Guillermo e eu, cal dos dous ten o dereito. [CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]
3. O Señor decidirá agora entre Guillermo e eu foi a súa resposta. [CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]
4. Berta: Nas relacións que houbo entre o Totó e eu, a familia estábache por iso máis presente do que nós mesmo pensabamos. [CORGA: A. R. Ballesteros, *Xogos de damas*, Xerais, 1988]
5. Fixo un aceno curioso que eu non souben interpretar e que, acaso, parecía pór en dúbida a miña madurez persoal ou, pola contra, acender unha faísca de alegría no seu interior, como se a

- constatación daquel mutuo interese entre Águeda e eu lle trouxese lembranzas doutros tempos, tempos nos que había unha muller na súa vida: a miña nai. [CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]
6. Hortensia: Entre eles e eu non lle hai ningunha diferenza. [CORGA: R. Vidal Bolaño, *Saxo tenor (Desgracia arrabaldeira improbable entre loucos, chourizos, gánsters, chulos, currantes e putas)*, Xerais, 1993]
7. Luís veuse sentar entre Gabriela e eu, quedouse mirando para as cubertas do libro e dixo: [CORGA: M. A. Murado, *Memoria de derribos*, Galaxia, 1993]
8. Non había, pois, nin un só punto de contacto entre S. Curtis Estelo e eu. [CORGA: X. Miranda, *A biblioteca da iguana*, Xerais, 1994]
9. Que diferenza con aqueloutra vez, a primeira, en que aqueles delicados, miúdos e feitucos se apertaban contra min enriba da herba, que diferenza, a pesar de que entre eles e eu se interpoñían as súas medias e o meu calcetín. [CORGA: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]
10. É triste que eu o diga pero a diferenza entre esas nenas e eu é mínima. [CORGA: N. Franco Barreiro, *Historias de Edel*, Edicións do Cumio, 1994]
11. Ó vela desde lonxe paréceme que foi cando por primeira vez me decatei das distancias que había entre ela e eu, coma da terra ó ceu, e sentín unha dor tan grande, e unha vergoña, e tamén unha alegría de contemplala alí e saber que, tanto coma ela valía, tivérame por seu amigo e acompañante, e mesmo me convidara dunha vez a tomar café no salón do pazo, coma un igual seu, que non fun capaz de achegarme a saudala para dicirlle que estaba preto súa. [CORGA: X. M. Martínez Oca, *Náufragos en terra. Os amores imposibles*, Xerais, 1995]
12. Entre García e eu hai unha brétema como a que se levanta do mar ás mañás, e a través dela repite, marche, Souto, esqueza a casa, non fedelle nesas historias, só os mortos poden tomar a requista ós mortos. [CORGA: M. Aleixandre, *Lobos nas illas*, Xerais, 1996]
13. É que a reacción entre ti e eu cando estamos a menos dun metro fai case cambiar a cor visible do aire, como ese licor anil do laboratorio que vira a vermello en presenza de azucre. [CORGA: M. Aleixandre, *Lobos nas illas*, Xerais, 1996]
14. Penso que foi un meniño entre ti e eu, debeu ser unha pulmonía ou así. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]
15. Pois entre Manuel e eu houbo un neno que polo visto morreu de meses e que se chamou Miguel. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]
16. Isto queda entre ti e eu, ¿está clariño? [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]
17. _Este vai ser un man a man entre ti e eu. [CORGA: F. A. Vidal, *A derradeira decisión*, Toxosoutos, 2001]

18. Resulta que despois daquela relación que aquela noite tivera coa bruxa, a pesar do satisfactoria que fora, medín ben as distancias entre ela e eu, polo que fose de caso, que con este xénero de seres sempre anda un xogado. [CORGA: I. Novo, *Airadas de palabras*, Laivento, 1998]
19. De súpeto deu un paso atrás e, entre el e eu apareceu o preso. [CORGA: X. Franco González, *Matalote*, Ir Indo, 2000]
20. Moza: Dende aí, dende o alto, has ter visto mellor do ca ninguén que malia o que sostivemos diante da xente, nada se consumou entre ese Vizconde e eu, fóra dunha promesa que agardo non ter que cumprir nunca. [CORGA: R. Vidal Bolaño, *A burla do galo ou Vida e amores de Don Esmeraldino da Cámara Mello de Lima Vizconde de Ribeirinha e Galo de Portugal*, IGAEM, 2000]
21. Coa base que entre a familia e eu logramos cimentar, despegou e foi aterrar a Nova York. [CORGA: CG2000-11-25/41]
22. Tardei en decatarme moito tempo de que el só era un mediador entre as miñas pantasma e eu. [CORGA: M. Romero, *Dubido se matei a Lena*, Xerais, 2000]
23. Que entre Raúl e eu o que houbo sempre non sei se foi amor, pero si que foi adicción, unha droga, o sabor do ron que se proba e logo non se pode deixar, ou que se deixa de mentira, de teatro, que sempre volvemos probalo, e sempre volvemos emborracharnos. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]
24. e malia ser sen estar nunca, Santi instalouse entre Raúl e eu, incluso despois de marchar e volverme deixar ao servizo de Raúl e o seu harén, enterrada en San Eirís, soa, pero coa certeza de seguir tendo dentro a alegría, a tristeza, o medo, a cólera, o erotismo e a tenrura, e por riba de todas, a sorpresa. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]
25. E ben pensado moi hixiénico non era, pero con elas facía patente a distancia entre eles e eu. [CORGA: M. Regueira, *L'affiche rouge*, Xerais, 2007]
26. Entre Andrés e eu, que eramos os máis vellos dos seis irmáns, comezamos a envasar coma auténticos lobos de mar e coas mesmas xeituras coma se estivésemos nos mares de fóra envasando para as lanchas grandes dos tarrafeiros. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

C) Ocorrencias de complemento con *min* como primeiro elemento da coordinación:

1. Sinto que a corrente de simpatía entre min e mailos meus alumnos tense tamén endurecido. [CORGA: X. L. Méndez Ferrín, *Arraianos*, Xerais, 1991]
2. Mal que ben, existe unha barreira defensiva, a barra, que se ergue entre min (ou nós, cando non estou só) e os da outra banda, porque de non ser así hai tempo que estaría vencido, dominado. [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

3. Entre min e mailo home que termaba do meu brazo. [CORGA: X. López López, *O caderno*, Sotelo Blanco, 2001]

D) Ocorrencias de complemento con *min* como segundo elemento da coordinación:

1. Saín da casa e botei a correr monte arriba, sempre monte arriba, como fago decote dende hai tempo; o mesmo tempo que hai dende que me dedico ó ioga e frecuente esotéricas relixións e ritos, abraiantes cultos de latría e hiperdulía, en alporizada e continua contradición entre eles e min mesma. [CORGA: A. Conde, *Memoria de Noa*, Xerais, 1982]
2. Eu xúrolle, Amaro, con tódalas miñas forzas, que quede ben claro, que entre Dolores e min non hai a mais mínima intimidade. [CORGA: Ú. Heinze, *O soño perdido de Elvira M.*, Xerais, 1982]
3. O buraco entre ti e min aumentaba sen remedio. [CORGA: R. Raña, *O crime da rúa da moeda vella*, Xerais, 1993]
4. Nin sequera eramos quen de cruzarmos unha palabra entre miña nai e min. [CORGA: L. C. Carballal, *Historias non irreais*, Espiral Maior, 2001]

Curiosamente, o castelán prescribe as formas de nominativo e relega as de complemento a épocas pasadas (RAE, 2005: 263), malia que no CREA se documentan algúns casos relativamente recentes con forma preposicional:

Mi hijo se bajó y la Eréndira se sentó entre mí y Rodolfo. [CREA: J. López Páez, *Doña Herlinda y su hijo y otros hijos*, Fondo de Cultura Económica, 1993]

"Yo estoy admirablemente ahora, como nunca, nunca en Santander, ni tampoco aquí al principio, nunca creí que alcanzaría un momento como éste, tan de unificación de mí misma, tan alegre, me siento muy alegre y cerca de todo lo cercano y lo lejano, tengo la impresión de que no existe la menor distancia entre mí misma, mis sentimientos, mi bienestar y el mundo. [CREA: Á. Pombo, *Una ventana al norte*, Anagrama, 2004]

No portugués, pola contra, parece que comeza a desenvolverse o mesmo proceso que xa está avanzado en galego: a substitución das formas de complemento polas rectas, posto que, a pesar de que se aconsella o emprego das formas oblicuas tras *entre*, “*na linguagem coloquial predomina, porém, a construção com as formas rectas, construção que se vai insinuando na linguagem literária*” (Cunha e Cintra, 1984: 301).

Os corpus do portugués, con todo, de momento apenas amosan ocorrencias, xa que, por exemplo, o *Corpus do Português* de Davies e Ferreira, fronte a catro casos coa

forma recta *eu*, achega 114 coa oblicua *mim*; mentres que o CRPC recolle dous coa recta e 318 coa de complemento¹⁷⁶. Vexamos un par de exemplos de cada un deles:

"Até hoje enviam passagens de Salvador para a minha casa, tentando estabelecer um vínculo entre eu e o Bahia". [Corpus do Português: *Tribunal julga na próxima semana processo contra juiz*, (4/3/1997)]

Entre mim e o Eduardo -olha que grande coisa!- há só dois anos de diferença. [Corpus do Português: F. Costa, *Carcere Invisível*, 1972]

Meus irmãos, diante de desafios de tamanha monta, Nino engana-se ao encarar estas eleições como se fossem uma luta mortal entre eu e ele, numa vã tentativa de se vingar da história. [CRPC]

"E o ponto comum entre mim e Bernard é esse: gostamos ambos da vida, continuamos o combate". [CRPC]

Álvarez (1999) aborda o estudo do estatus das formas pronominais tónicas da primeira persoa coa preposición *entre* atendendo aos datos históricos, á situación no portugués moderno e mais aos datos orais e literarios do galego contemporáneo, e defende o proceso substitutorio de *min* por *eu* non como un castelanismo senón como un fenómeno que afecta en maior ou menor medida a todos os romances peninsulares. A autora, así mesmo, defende a norma actual para o significado de coparticipación baseándose na equivalencia semántica co suxeito e na “*analogía con frases compostas que cumpren esta función* (facémolo entre eu e el, *coma* facémolo eu e el)” (1999: 138); finalmente, indica que a substitución de *min* por *eu* vai máis retrasada con outras funcións, polo que aínda que non deben condenarse con eses outros significados as formas de nominativo si deben prescribírense as de complemento.

Como acabamos de comprobar, a lingua non é algo estático, senón que está con cambios constantes, máis ou menos perceptibles; somos actores e espectadores dun cambio lingüístico en concreto no que nunca reparáramos. Mais, como afecta este proceso á etiquetaxe do noso corpus? Que criterio utilizar á hora de caracterizar os pronomes tónicos nos que se dá sincretismo entre o caso nominativo e o complemento cando seguen á preposición *entre*? Difícil decisión, mais hai que tomar unha. Por unha banda, fundamentar a distinción entre forma recta e oblicua tras *entre* baseándonos nos

¹⁷⁶ As consultas realizadas [10/05/2011] en ambos os dous corpus foron as secuencias “entre eu e” e “entre mim e”, co que case nos aseguramos de que o termo da frase preposicional sexa coordinado.

diferentes significados que achega a preposición non é aconsellable en canto que observamos no CORGA, malia que sexan os menos, casos de localización con forma nominativa e casos de cooperación/coparticipación con forma preposicional:

De súpeto deu un paso atrás e, entre el e eu apareceu o preso. [CORGA: X. Franco González, *Matalote*, Ir Indo, 2000]

Houbo un cativo forcexo entre nolos dous. Entre min e mailo home que termaba do meu brazo. [CORGA: X. López López, *O caderno*, Sotelo Blanco, 2001]

Ademais, na determinación das categorías e atributos prescindimos sempre da semántica, polo que sería incongruente recorrer a ela agora; por riba, prevendo as dificultades, é máis que probable que necesitésemos implementar unha regra lingüística para axudarlle ao etiquetador a decidir cando corresponde cada un dos valores, labor este que, nun sistema no que impera a morfoloxía e os aspectos semánticos brillan pola súa ausencia, non nos vemos quen de realizar, amén da complexidade que suporía informaticamente implementala.

A opción pola que nos decidimos no CORGAetq é sinxela, aínda que non por iso quizais a máis acertada, e consiste en caracterizar o caso nominativo tras *entre* só cando temos unha forma recta de primeira persoa singular, ou de segunda se no texto se mantén a oposición *tu/ti*, mentres que no resto dos casos marcamos o caso preposicional, caso que por regra, coa excepción das modais e excluíntes, lles corresponde aos pronomes que son termo de frase preposicional.

Finalmente, non queremos concluír esta epígrafe sen comentar que, á hora de decidir o lema que lle corresponde a cada unha das formas do paradigma pronominal, optamos por manter na lematización todas as diferenzas, excepto as de xénero e número. Así, os lemas que consideramos son os seguintes:

serie tónica: *eu, min, comigo, ti, contigo, vostede, el, il, si, consigo, nós, nosoutros, connosco, vós, vosoutros, convosco*;

serie átona: *me, che, te, lle, o, se, nos, vos*.

En realidade, se elimináramos as variacións de xénero, número e caso, non deberían existir máis ca tres lemas na clase Pronome: *eu, ti* e *el*. Non obstante, a lematización que propoñemos facilita estudos moi refinados do uso dos pronomes e, ao

tempo, non empece en ningún momento un estudo máis xeral, xa que as buscas por etiqueta permiten tamén esta última opción.

4.4.2. *Un, unha, uns, unhas*

Malia que a adscrición de *un* ás categorías gramaticais nas que tradicionalmente se inclúe é máis problemática do que nun principio puidera pensarse¹⁷⁷, para a súa

¹⁷⁷ Na tradición hispánica a caracterización categorial de *un* foi e segue sendo obxecto de debate (RAE, 2009: 1090-1092). Así, por exemplo, Alarcos sinala:

La distinción tradicional entre uno numeral, uno pronombre indefinido y un, una, unos, unas como artículo indeterminado carece de justificación. Su comportamiento funcional es unitario y la referencia que efectúan análoga. No puede ser artículo por cuanto este (el, la, etc., §79) carece de acento y no es palabra independiente. Separar, de otra parte, el sentido numeral respecto del indefinido es innecesario, ya que de todas las maneras se trata de un cuantificador. [Alarcos, 1994,:122]

A tradición galega non é allea a esta controvertida clasificación. Se revisamos as gramáticas e manuais máis recentes comprobaremos que, malia a pretensión de delimitar *un* en tres clases ou subclases categoriais diferentes, o establecemento dos límites entre o tradicional artigo indeterminado e o cuantificador indefinido non están nada claros. Así:

- Carballo, no apartado dos indefinidos (1976: 197) apunta que “Un, unha, uns, unhas *en función adjectiva se confunde con el artículo indeterminado. Como pronombre sustantivo, carece de plural*”.
- Costa Casas *et al.* (1988: 121-126 e 137-139) non consideran a existencia da clase Artigo, senón de Pronome, e dentro desta clase inclúen:
 - a subclase Identificadores, que acolle tanto as formas do que nós denominamos artigo determinado (*o, a, os, as*) como as do artigo indeterminado (*un, unha, uns, unhas*) e o chamado pronome indefinido en plural (*uns, unhas*); pero iso si, *o (a, os, as)* sempre ten función adxectiva ou determinante, mentres que *un (unha, uns, unhas)* tanto a ten adxectiva como substantiva ou nuclear (*Por aí pasaron uns colegas / por aí pasaron uns*).
 - a subclase Cuantificadores, que no subtipo indefinidos inclúe os plurais *uns, unhas* para os que se especifica que non teñen singular, porque senón serían cuantificadores unitarios [= numerais] ou identificadores indefinidos.

Nós preguntámonos cal é a diferenza entre os exemplos achegados, na súa terminoloxía, para o identificador indefinido “*Trouxéron-mo uns rapaces*” fronte ao do cuantificador indefinido “*Chegaron uns vehículos motorizados*”, pois non conseguimos apreciála.

- Nas *Normas* (RAG/ILG, 1985: 100; 2003: 99), á parte do recoñecemento da categoría Artigo indeterminado, despréndese a consideración de *un(s), unha(s)* tamén como Indefinido, tanto con función substantiva como adxectiva: *falouse duns problemas que non demos entendido; cuns veciños podemos falar, con outros é imposible*”, mais en ningún momento se establecen os límites entre as formas do artigo indeterminado e estas do indefinido.
- Freixeiro (2000: 250, 261, 287), a semellanza do que, *grosso modo*, fan Costa Casas *et al.*, integra todos os usos de *un, unha, uns, unhas* nos pronomes, diferenciando entre:
 - identificador indefinido (algúns dos usos dos tradicionais artigo indeterminado e do pronome indefinido): *Chegou un amigo seu; uns dálle por aquí e outros por ali...*;
 - cuantificador indefinido (só en plural: *uns, unhas*) con valor equivalente a *algúns*: *Levaba unhas coartelas no peto da chaqueta*;
 - cuantificador definido (o tradicional numeral, *un, unha*): *Coustome unha peseta*.

diferenciación nós baseámonos na *Gramática Galega* de Álvarez, Regueira e Monteagudo (1986: 140-141; 155-156; 232-234), segundo a cal:

A) *un, unha, uns, unhas* considerámoslos elementos da categoría Artigo, *tipo indeterminado*, cando funcionan como determinantes, isto é, cando acompañan un substantivo co que conforman unha unidade. Exemplos:

O Fondo Ivercaixa Galicia recibirá unha achega inicial de 5.000 millóns de pesetas para favorece-lo desenvolvemento de novos proxectos empresariais. [CORGAetq: CG2000-12-19/12]

A ausencia de Touriño converteu nun acareo Quintana-Feijoo 'Hai debate!' [CORGAetq: DLV2007-05-15/17]

Así, en Infantil matricularanse uns 22.000 nenos, que abranguen unhas 14.000 familias; uns 2.140 en Primaria, dunhas 1.430 familias; uns 13.260 niso, dunhas 8.800 familias; e uns 5.110 en Bacharelato, de 3.400 familias. [CORGAetq: GH2007-03-17/29]

Con todo, non acaban de aclararse os criterios que permiten discriminar os usos do cuantificador indefinido dos do identificador indefinido, porque na nosa opinión *Uns dálle por aquí e outros por alí* ademais de identificar tamén cuantifica, e a conmutación por *algúns* é válida en ambos os casos.

- Álvarez e Xove (2002: 454-457) consideran que só os plurais *uns, unhas* son cuantificadores non comparativos de non totalidade non paramétricos, o que vén significar que son cuantificadores que non expresan determinación cuantitativa respecto dun termo de comparación expreso ou identificable no contexto, é dicir, que non son *máis, menos, tanto (tantos, tanta(s))*, que expresan unha cantidade relativa (oposta á absoluta expresada por *todo* ou *ningún*) que en xeral é pequena e que non se establece respecto dun parámetro. Para a súa diferenciación co artigo indeterminado -ao que tamén lle conceden función nuclear- remiten ao apartado 5.5.1.3.2, mais nós non o atopamos por ningures, co que nos queda sen aclarar en que fundamentan a distinción, xa que na epígrafe do artigo indeterminado só se indica que as formas *uns, unhas* poden empregarse como cuantificadores indefinidos co significado de cantidade pequena ao que vai unido en xeral o significado específico de artigo indeterminado, mais que non necesariamente implica cantidade pequena (*sic*).
- Xove (2010: 247-261) volve insistir na existencia do artigo *un, unha, uns, unhas* como núcleo dunha frase substantiva e invita a que se corrixa as actuais *Normas* (RAG/ILG, 2003) baseándose en que o cuantificador é sempre plural e en que os exemplos do “pronomo cuantificador” en realidade sono de “indefinido identificador”, por canto neles non se cuantifica nada. Non obstante, a pesar de revisar a clasificación de *un* na historia recente das gramáticas e manuais do galego en ningún momento dá conta dos criterios nos que se fundamenta a diferenciación que el propón nin explica cando consideralo artigo e cando cuantificador.
- Hermida (2004: 45-55) obvia toda a problemática ao non tratar a existencia de *un* entre os indefinidos, non sabemos se debido a un lapsus ou non, malia que logo cando define o cuantificador *canto(s), canta(s)*, avisa de que “*non se usa de non ser precedido polo indefinido un, unha, uns, unhas (baixei uns cantos quilos)*”.
- López Viñas et al. (2010), finalmente, clasifican o Artigo (determinado e indeterminado) nunha subclase do Pronome e caracterízano como anteposto sempre a un substantivo, coa opción de concorrer co posesivo, cuxo valor é o de actualizador; non obstante, cando o describen concédennlle tamén a función de Núcleo da frase substantiva (p. 268). Así mesmo, á parte do Numeral *un, unha*, consideran tamén a existencia de *un, unha, uns, unhas* nos Cuantificadores, pero, a pesar de incluílos na táboa dos cuantificadores, logo non os desagregan nin describen, razón pola que seguimos ignorando onde radica a diferenza entre as formas do artigo e as do cuantificador.

No continente é improbable un Bush, unha Thatcher a pesar das ensoñacións dos intelectuais liberais. [CORGAetq: *TN1998-05/18*]

B) *un, unha* caracterizámolos como numerais unicamente nunha relación numérica ou cando o contexto nos indica de maneira clara que estamos ante un numeral. Así pois, seguido de substantivo co que forma unha unidade, o artigo será o termo non marcado da oposición categorial Artigo / Numeral. Vexamos algunhas ocorrencias da clase numeral:

O pasado agosto, o paro rexistrado na zona euro (os países da moeda única) situábase no 9,1 por cento, unha décima por encima do índice actual, e en setembro do 1999 nun 9,8 por cento. [CORGAetq: *CG2000-11-08/90*]

Aos Socialistas aínda lle falla perder unhas cantas votacións para entender que as opcións dos seus socios potenciais xa non son ou co goberno ou coa abstención, así que quixeron matar dous paxaros dun tiro: á comisión e a Quintana. [CORGAetq: *ANT2008-07-10/20*]

Na operación inclúese unha oficina do Urquijo en León, tres oficinas en Xixón, unha en Avilés e unha en Palencia, todas estas dependentes da dirección de zona de Caixa Galicia que ten a súa sede na localidade de León. [CORGAetq: *CG1996-02-10/44*]

cinco en 2005 e un no período de 2004, segundo os datos facilitados polo Centro de Xestión de Tráfico de Galicia. [CORGAetq: *GH2008-08-05/24*]

C) *un, unha, uns, unhas* analizámolos como elementos da clase Indefinido cando non actúan como determinantes, é dicir, cando son núcleos dunha frase substantiva. O mesmo, para as formas do singular, se non sinalan cantidade. Exemplos:

Uns porque ven que a economía empeza a estar supeditada ó que dean de sí os fluxos informativos, por pura especulación outros, e hainos que entran nese mundo coa secreta intención de acumular poder persoal. [CORGAetq: *GI1996-06/31*]

"Nin o portugués nin o castelán deben ser prioridades, pero está claro que hai obras fundamentais que deben ter a súa versión galega e, inequivocamente, o Quixote é unha delas. [CORGAetq: *ANT2005-03-03/33*]

Alí puido debutar co primeiro equipo coruñés e coñecer a Andrade, un dos seus ídolos. [CORGAetq: *DLV2007-05-15/110*]

Nas cidades hai diferenzas entre unhas e outras, por exemplo de Santiago a Vigo ou A Coruña, e temos que traballar máis porque é onde se concentra máis poboación. [CORGAetq: *ANT2008-07-10/15*]

Con todo, é previsible que o interese por facer o que a un lle gusta ocupe un lugar privilexiado tanto para os pequenos como para os adolescentes: [CORGAetq: GH2008-08-05/29]

4.4.3. Como

Sen dúbida ningunha, xunto con *que*, *como* é unha das formas gráficas máis frecuentes do corpus *-que* ocupa a segunda posición con 979 763 ocorrencias e *como* a vixésima primeira con 129 223 aparicións¹⁷⁸, pero tamén se trata dunha das máis complicadas para etiquetar.

Moitas son as páxinas que sobre esta unidade se teñen escrito, máis nós non imos esquematizalas aquí nin tampouco centrarnos na súa análise. A nosa pretensión é só tentar explicar que caracterizacións lle atribuímos nas diferentes ocorrencias do CORGA etiquetado.

Comezaremos por lembrar que a unidade ortográfica *como*, alén da primeira persoa do presente de indicativo do verbo *comer* e da súa inclusión nunhas cantas locucións¹⁷⁹, é etiquetada no noso sistema nuns casos como conxunción e noutros como adverbio, ben interrogativo-exclamativo ben relativo.

É verdade que nos últimos anos, tanto na tradición galega (Álvarez e Xove, 2002) como na hispánica (RAE, 2005), téndense a clasificar certos usos deste *token* na categoría preposicional, mais ningún dos manuais que seguimos lles atribúe esa función¹⁸⁰, motivo polo que decidimos non integrar *como* na relación de preposicións e, por eliminación, seguir remitindo esas ocorrencias ao adverbio relativo. Estamos referíndonos a exemplos coma os seguintes, todos eles caracterizados como adverbios relativos e non obstante susceptibles de ser conmutados por unha preposición:

Estas entidades actuarán como aseguradoras da operación, mentres que o Instituto de Crédito Oficial (ICO) o fará como axente financeiro. [CORGAetq: CG1996-02-10/100]

A principios dos 90 saía en Francia un libro de Michel Albert, Capitalismo contra capitalismo, que constituía unha importante contribución á antroploxía económica e como tal foi acollido pola crítica máis atenta. [CORGAetq: TN1998-05/18]

¹⁷⁸ Datos extraídos da versión 1.5 das frecuencias descargables do CORGA.

¹⁷⁹ Entre elas: *así como*, *ben como*, *mal como*, *como queira que*, *como mínimo*, *como máximo*, *como moito*, *como pouco*.

¹⁸⁰ Non aparece na listaxe de preposicións das *Normas* (RAG/ILG, 1995, 2003) nin en Álvarez, Regueira e Monteagudo (1986), obras ambas nas que fundamentamos no sistema XIADA o establecemento do etiquetario e a etiquetaxe posterior.

Viñeron falar como afectados en primeira persoa do sinistro, e só houbo dous comparecentes que recoñeceron as perdas produtivas como consecuencia do Prestige. [CORGAetq: ANT2005-03-03/22]

Pola súa banda, o coordinador xeral de IU, Gaspar Llamazares, acusou a Díaz de Mera de actuar como o "peón negro" na teoría da conspiración posta en marcha polo PP, hipótese, dixo, que "se desmorona como un castelo de naipes". [CORGAetq: GH2007-03-30/55]

Outro dos usos preposicionais máis habituais de *como*, que no noso sistema hai que buscar entre os asignados ao adverbio relativo, é o de introdutor dun Predicativo, mais, a diferenza dos inmediatamente anteriores, nestas circunstancias a súa presenza acostuma ser opcional:

Malia a dimensión xa alcanzada, o declive económico e o deterioro social que padece o noso País maniféstase como unha aberración que é posible superar. [CORGAetq: TN1997-05/1]

O mercado de bonos, por outra parte, percíbese como perdedor no caso de que o gobernador de Texas chegue á Casa Branca xa que se tería que cumpri-la súa promesa de reduci-los impostos, reducindo así o superávit do goberno federal. [CORGAetq: CG2000-11-10/77]

No Parlamento Español, tanto o deputado nacionalista Francisco Rodríguez, como o socialista Ceferino Díaz, criticaron duramente o que consideran como "orçamento enganoso e insuficiente". [CORGAetq: ANT2002-10-03/31]

Sen o menor atisbo de imposición, podemos considerar como territorios próximos a Brasil, Cabo Verde e por descontado Portugal, pois foi en Sintra _localidade próxima a Lisboa_ onde se materializou a gravación que se nos ofrece. [CORGAetq: ANT2008-07-10/48]

'Defínome como federalista'. [CORGAetq: ANT2008-07-10/22]

Dadas as dificultades que entraña, pola falta de fronteiras nidas, discriminar entre os usos do *como* conxunción subordinante modal e o *como* relativo, e posto que a gramática de Álvarez, Regueira e Monteagudo (1986: 285-286, 548), obra na que fundamentamos a etiquetaxe do noso sistema, o consideran sempre adverbio, tamén no sistema de XIADA ambas as ocorrencias van da man, caracterizándoas sempre como adverbio relativo. Somos conscientes dos obstáculos que esta decisión implica á hora de realizar un estudo sobre esta forma en particular, mais tampouco debemos perder de vista que 1) o obxectivo principal do corpus é servir de adestramento para un etiquetador automático estatístico, o cal vai inferir as regras de futuras etiquetaxes precisamente a partir de múltiples oracións etiquetadas manualmente; 2) a coherencia é requisito imprescindible para esta ferramenta de análise e 3), se os lingüistas non nos

poñemos de acordo en que é relativo e que é conxunción modal, moito menos seremos capaces de implementalo con regras ou de levalo á practica nunha etiquetaxe congruente.

Unha pequena mostra desta problemática amósanola a suposta obrigatoriedade que algúns autores, incluídos entre eles os que nós seguimos, lle outorgan á presenza do antecedente para considerar relativo *como*:

O peor de todo foi o xeito como llo dixo (Álvarez, Regueira e Monteagudo, 1986)

Molestoume a maneira como trataba ós rapaces (Álvarez e Xove, 2002)

O xeito como llo dixo non lle gustou nada (Hermida, 2004)

Na nosa opinión, *como* xa posúe no seu contido a noción léxica de modo, igual que os outros relativos *cando* e *onde* a posúen respectivamente de tempo e lugar. Así é que, como falantes, os exemplos anteriores resúltannos forzados, precisamente por estar explícito o antecedente, xa que apreciamos máis naturalidade nas secuencias:

O peor de todo foi como llo dixo

Molestoume como trataba ós rapaces

Non lle gustou nada como llo dixo

Unha vez máis na nosa opinión, o feito de que estea explícito o antecedente o único que fai é favorecer a conmutación do relativo, sexa *como*, *cando* ou *onde*, por unha frase preposicional co relativo máis funcional, *que*:

O peor de todo foi o xeito en que llo dixo

Molestoume a maneira en que como trataba ós rapaces

O xeito en que llo dixo non lle gustou nada

Evidentemente, necesitaríamos profundar no estudo desta partícula, mais *a priori* parece que nos textos do CORGA, referencia do uso do galego actual, prefírese prescindir do antecedente e limitarse á utilización de *como*. As consultas dun termo léxico xenérico que indique modo, coma *xeito*, *modo*, *maneira*, *forma*, etc., seguido de *como* amosan resultados escasos e non obstante estamos seguros de que unha alta porcentaxe das case 130 000 ocorrencias do corpus corresponden á análise relativa, pois cal é, senón a explicitación do antecedente, a diferenza entre as secuencias de a) e as de b)?

a)

A Pena parece importarlle pouco ademáis esto, a xulgar pola maneira como resolvéu a cuestión polo de agora. [CORGA: TE1977-13/1]

Sobre esta pronunciou espectaculares conferencias no Círculo de las Artes con auténtico abarrote de público, un público verdadeiramente hipnotizado non polo que dicía senón pola forma como o dicía: [CORGA: M. A. Murado, *Memoria de derribos*, Galaxia, 1993]

Este foi o xeito como Nattfari se procurou o carbón que precisaba. [CORGA: X. Bernárdez Vilar, *A saga da illa sen noite*, Toxosoutos, 2001]

De cada arte dáse unha explicación dos elementos que a compoñen, da clase á que pertencen, da súa forma, do modo como se emprega, dos peixes que con ela se collen e da súa evolución histórica. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

b)

Cando me venceu a curiosidade fun testemuña xustamente de como se fundían cara a abaixo os banzos superiores da escaleira, coma un bulto deformado. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Nada caros, xulgando por como concluían a descrición do local os que pretendían orientar ós forasteiros. CORGA: X. González Martín, *Vigo. No espello dos nosos avós*, Xerais, 1991]

Un capataz que traballou para a Deputación de Lugo nas estradas comenta que o problema non só "está na calidade dos tubos, senón en como se asentan". [CORGA: ANT2006-12-21/24]

Carta que imos reproducir nalgúns fragmentos non só polo que di, senón por como o di. [CORGA: ANT2007-02-22/24]

A non diferenciación no noso sistema entre os usos preposicionais, os conxuntivos modais e os relativos provoca que documentemos na versión 2.4 do CORGA etiquetado 783 casos de *como* relativo, encontrándose entre eles ocorrencias dos tres tipos anteriores. Vexamos unha pequena mostra:

As cousas son como son.[CORGAetq: GI1996-06/30]

Así foi como comezaron as reestructuracións nos estaleiros, sempre coñecidas como reconversións. [CORGAetq: TN1997-04/7]

"Logo de meses de preparación dun proxecto, non podemos estar pendentes dunha axuda de Industria que ó final se nos vai denegar, como xa sucedeu co proxecto Textilive, por razóns políticas", conclúe Mar Iglesias. [CORGAetq: TN1997-04/7]

Os Quince non falaron de cifras concretas, pero como afirmou o vicepresidente do Goberno e ministro de Economía, Rodrigo Rato, a UE "será, como sucedeu en Bosnia, o contribuínte máis importante". [CORGAetq: CG1999-05-11/18]

Tal e como recorda Francisco Rodríguez, "estes cartos non van quedar en Galicia, como si acontece cando se fai unha autovía ou un aeroporto. [CORGAetq: ANT2002-10-03/31]

O documento fala en voz alta do pobo galego, dándolle o papel protagonista no Preámbulo do Estatuto, tal como correspondería a un texto de carácter constitucional. [CORGAetq: ANT2005-03-03/5]

"A consolidación do sector forestal galego como actividade economicamente competitiva pasa pola investigación e a correspondente aplicación práctica como instrumento de crecemento e competitividade", indica o estudo. [CORGAetq: ANT2005-03-03/56]

"A calquera persoa que resida nalgunha das cidades galegas, vote como vote e pense como pense, gustaríalle que a súa cidade fose como Pontevedra, e por iso os pontevedreses séntense orgullosos da súa cidade". [CORGAetq: GH2007-03-17/20]

Eliminado o uso modal de *como* das conxuncións e dado que nestas non se distinguen subtipos en XIADA, senón que a diferenciación soamente se realiza entre coordinantes e subordinantes¹⁸¹, os diferentes valores semánticos que ofrece o *como* conxunción agrupámoslos na etiqueta Cs: *Conxunción subordinante*. Así, os usos propiamente comparativos:

_Esas persoas que foron de maneira moi premeditada ó negocio da comunicación enturviaron, desgraciadamente, algo tan noble como é que unha entidade financeira poida participar nunha empresa xornalística, ou nunha empresa de cable. [CORGAetq: GII996-06/31]

Na práctica resulta difícilísimo conseguir a través da Administración galega algo tan básico como un simple formulario de solicitude". [CORGAetq: TN1997-04/7]

Entrou como un canón con Tesis. [CORGAetq: ANT2005-03-03/18]

Se alomenos gaña tanto como perde, se engade algún aspecto novo, entón si que a fago. [CORGAetq: ANT2005-03-03/49]

O primeiro que te atopas ao entrar son os nosos maiores pasando o tempo entre as cartas, a conversa amable e as lecturas de xornais como este. [CORGAetq: GH2007-03-30/112]

os causais:

¹⁸¹ Véxase o apartado relativo á categoría Conxunción no Capítulo I.

Como estou casado cunha valenciana, fago ás veces o percorrido desde Valencia a Galicia, e teño que dicir que me sinto identificado con tódolos pobos polos que paso. [CORGAetq: *GI1996-06/30*]

"O médico díxome que debín de perder o DIU, que como é unha cousa pequena puiden perdelo polo servizo sen decatarme", recordou. [CORGAetq: *GH2007-03-17/41*]

Como está nun instituto podemos observar as obras de arte mentres os rapaces veñen e van, con apuntamentos, libros, mochilas. [CORGAetq: *GH2007-03-30/109*]

Como é seguro que as pistas non abundan, diremos que estamos a falar de Raquel Welch, quen tomou o apelido do seu primeiro marido, James Welch. [CORGAetq: *ANT2008-07-10/54*]

Como aborrezco o fútbol e os seus fastos rabañís, durante o último campionato só logrei tirar dúas conclusións que me resultasen proveitosas. [CORGAetq: *ANT2008-07-10/9*]

os completivos:

É certo que se miramos minimamente a historia, observaremos como hai na península ibérica un decreto de nova planta no século XVIII que é determinante para o desenvolvemento de determinadas rexións. [CORGAetq: *GI1996-06/30*]

O banco viu cómo o seu beneficio neto no primeiro trimestre se reducía en máis dun 78% con respecto ós 481,8 millóns de euros conseguidos nese período do 2001. [CORGAetq: *CG2002-05-20/9*]

Con Soños tripados, Muras dá a coñecer como "cos pés enriba da terra se pode soñar o que un queira para este país, para esta terra, para este planeta". [CORGAetq: *GH2007-03-17/12*]

Non nos queremos dar conta do sufrimento que tiveron moitas xeracións anteriores que vían como perdían os fillos, como desaparecían familias enteiras. [CORGAetq: *ANT2008-07-10/11*]

Eu vexo como moitos rapaces sofren na clase cando se empregan encerados electrónicos. [CORGAetq: *ANT2008-07-10/50*]

ou os condicionais:

Creo que non... como non o remediemos. [CORGAetq: *MC1980-12/1*]

Debido á súa proximidade cos usos comparativos, caracterizamos tamén como conxunción subordinante o emprego de *como* nas exemplificacións:

Países como Bélxica, Italia, Irlanda, Holanda ou Suecia terán unha débeda pública moi superior a ese 60 por cento do PIB, aínda que en vías de redución, polo que, para evitar que queden excluídos, é necesario que os ministros elaboren unha "interpretación transparente e obxectiva" do criterio. [CORGAetq: *CG1996-02-07/103*]

Vega destacou que en departamentos como Pesca as modificacións supoñen incrementar en máis dun 50 por cento as contas iniciais. [CORGAetq: *CG1999-05-22/69*]

Así mesmo, explicou que AA abrirá novas rutas en destinos como Cuba, Israel e zonas de Asia e Oceanía. [CORGAetq: *CG2001-10-23/44*]

Na maioría dos casos trátase de empresas con máis de 20 anos no sector como Afinsa, Forum Filatélico, ou mesmo Correos. [CORGAetq: *ANT2002-10-17/28*]

Ademais dos países avanzados, outros como Rusia, Cuba ou Venezuela, procederán axiña á substitución das antigas lámpadas de filamento polas de baixo consumo. [CORGAetq: *ANT2008-07-10/55*]

cando, seguido de *se*, establece unha comparación non real:

Non se deben analizar as perspectivas industriais como se continuara o auxe económico dos anos sesenta e non mediase a crise do sistema produtivo, comercial e financeiro mundial, nen razonar desde simples supostos económicos, cando nos atopamos nun periodo de transición no que está en xogo o autogoberno político, decisivo, no caso de se producir, no futuro da sociedade galega. [CORGAetq: *MC1980-12/1*]

A picaresca, sen embargo, multiplica os ibéricos de belota como se fosen fungos. [CORGAetq: *CG2001-10-10/35*]

Coido que estamos acostumbrados a ler toda a información como se se tratase dun texto literario. [CORGAetq: *ANT2005-03-03/71*]

Como se dun descenso consumado se tratase, as caras de Oubiña ou Gustavo López parecían asumir un fracaso que aínda é remediabile, malia que o equipo vigués depende agora dos erros do Athletic de Bilbao. [CORGAetq: *DLV2007-05-15/103*]

O grupo afrontou esta nova entrega como se fose o seu debut, dende o punto de vista da produción artística e executiva. [CORGAetq: *ANT2008-07-10/48*]

cando, seguido de *quen*, establece unha comparación falsa:

Estamos xa como quen di ás portas da entrada en vigor do euro como moeda única comunitaria e da desaparición da peseta. [CORGAetq: *CG2000-11-22/77*]

Logo, movendo así a man como quen non quere a cousa, dei por concluído o asunto. [CORGAetq: *Ú. Heinze, Remuíños en coiro, Xerais, 1984*]

ou cando, seguido de xerundio, establece unha comparación hipotética:

Ante esta cuestión, o axente encolleuse de ombreiros como dicindo "é o teu problema", e non contestou. [CORGAetq: *GH2007-03-17/40*]

Incluimos tamén na categoría Conxunción e no subtipo *subordinante*, as ocorrencias do *como* seguidas de “para + infinitivo” ou de “para que + subxuntivo”, que na maior parte das veces pode suprimirse sen que cambie o sentido do enunciado:

Non debemos ser tan ilusos, polo tanto, como para pensar que nos atopamos ante o inminente derrumbamento do inxusto e despilfarrador sistema económico que padecemos, e próximos á súa substitución por unha economía non dependente e socialista en Galiza, libre, por outra parte, dos múltiples erros e deformacións dos sistemas económicos que, deica hoxe, levan ese nome. [CORGAetq: MC1980-12/1]

Non son un home público como para andar tódolos días rondando a prensa. [CORGAetq: GI1996-06/31]

A Intertoto, e moito máis a UEFA, quedan xa a demasiados puntos como para pensar que o conxunto de Caparrós será capaz de chegar a Europa por algunha vía. [CORGAetq: DLV2007-05-15/105]

Si, aínda que se trata dun festival non tan masivo a nivel de actuacións como hai outros, neste caso hai cinco ou seis grupos por día, que xa é bastante, pero que non é un número tan grande como para que o público vaia esgotado nin para que os grupos teñamos cen problemas coas horas de proba, os retrasos e demais. [CORGAetq: GH2007-03-30/111]

Pero tampouco debemos ser tan inocentes como para pensar que en Galicia sabiamos o que non coñecían noutros lugares. [CORGAetq: GH2008-08-05/93]

Finalmente, polo que respecta aos usos conxuntivos, decidimos considerar tamén subordinante o seu valor tanto na locución *así como* como a súa aparición na locución discontinua *tanto... como*.

Hermida (2004) afirma que existe un nexo copulativo, restrinxido á terceira posibilidade, cun significado equivalente ao de *e* que é *así como*:

Non debe confundirse este nexo copulativo coa estrutura *así como/así coma* comparativa que aparece en oracións do tipo de *non sempre me visto así coma hoxe; así na terra coma no ceo*; nin coa estrutura *así como*, nexo temporal que indica simultaneidade na realización dunha acción: *así como veu, marchou*. [Hermida, 2004: 235-236]

Vexamos algúns exemplos extraídos da versión 2.4 do CORGA etiquetado:

O venres día 2, este saldo estaba composto por 4,23 billóns investidos polos estranxeiros en compras ata o vencemento de títulos como bonos e obrigas do Estado, así como por outros 42.300 millóns de letras do Tesouro (42.400 millóns ó día 26). [CORGAetq: CG1996-02-07/93]

A oposición parlamentaria foi exercida por unha estreita franxa da representación comunista e verde así como polos nacionalistas gaullistas franceses e pola extrema dereita. [CORGAetq: TN1998-05/18]

O incremento do prezo do petróleo rexistrado nos últimos meses debería presentar unha maior expansión do crecemento e das taxas de inflación dos países exportadores, así como unha apreciación real do tipo de cambio e un mellor comportamento do sector exterior. [CORGAetq: CG2000-11-06/54]

A escritora María Reimóndez e o fotógrafo Paco Vilabarro veñen de visitar Asia, onde observaron de perto os efectos do maremoto, así como o papel das organizacións de axuda internacional. [CORGAetq: ANT2005-03-03/66]

Serán diarias as viaxes que saen de Santiago cara ao aeroporto ás 10.00, 14.00 e 17.30 horas, así como as viaxes do aeroporto ao centro da cidade das 9.30 e 17.00 horas. [CORGAetq: DLV2007-05-15/14]

Os seus traballos sobre pensamento e filosofía, maiormente os dedicados á saudade, así como as súas traducións en colaboración con Celestino Fernández de la Vega das obras de Pokorny e Heidegger, supuxeron no seu momento un paso importante na normalización do galego como lingua con capacidade para expresarse en todos os campos do saber, superando a súa cualificación como lingua "rural e poética" na que a tiñan encadrada moitos intelectuais españois da posguerra. [CORGAetq: ANT2008-07-10/21]

As *Normas* (RAG/ILG, 1995 e 2003) e a *Gramática* de Álvarez, Regueira e Monteagudo (1986) non se pronuncian sobre a categorización de *así como*; é máis, nin o recollen. Na nosa opinión é incuestionable que se produce unha adición de elementos, e de aí o valor copulativo desta locución; mais acaso esa adición non se produce porque subxace nela tamén unha estrutura comparativa de igualdade?

En primeiro lugar, a nosa hipótese apóiase na concorrencia da variante *así coma* co mesmo significado copulativo ca *así como*. No corpus de adestramento, casualmente, non se atopan exemplos, mais no CORGA non é infrecuente a súa aparición nos distintos lustros nos que se clasifican os diferentes documentos aténdose ao período cronolóxico que ata o momento abrangue o corpus: 1975-2009. Vexamos unha pequena mostra representativa de cada período:

5.º A Xefatura Provincial de Sanidade realiza continuamente análises nos vertidos de Tafisa, Celulosas e Elnosa así coma nos peixes e mariscos da Ría e lamas da mesma. [CORGA: *Contaminación industrial e desenvolvemento. Mesa redonda encol de "Desenvolvemento económico e contaminación industrial na Galicia"*, SEPT, 1975]

reivindicábase o dereito do "pobo galego a auto-organizarse e conqueri-lo control dos seus medios de produción e destino", así coma "a participación popular na elaboración, control e administración da xestión pública" e propugnábase a falla dunha planificación que fose axeitada ós verdadeiros intereses globáis do pobo galego. [CORGA: TE1977-06/2]

Por suposto, non esquecemos que unha serie de labores así coma os propios animais, ademáis de facerse ou cobixarse nas plantas baixas, poden atopar o seu espacio en edificacións complementarias. [CORGA: F. Nebot, D. Pino e C. Portela; *O planeamento no mundo rural galego. Un exemplo: O Plan Xeral de Ordenación de As Neves*, Ediciones de la Torre, 1981]

Entre os narradores do que xa se ven chamando Xeración ou Grupo do Pedrón están Alcalá, Taibo, Martínez Oca, Rei Ballesteros, Tucho Calvo, Freixanes, Reigosa, etc., ós que habería que engadi-los nomes de A. Conde, Xesús Rábade, Helena Villar, entre outros, así coma o nome do non tan novo Riveiro Loureiro, accésit en varios dos Pedróns. [CORGA: F. Fernández Rei, *Prólogo a Nun cuarto para agardar / O último hexagrama / Cantar sen xesta de Lanzarote do Lago / Destino de Lumia / Grabación da volta a Galiza, dende o pazo de Raxoi*, Edicións do Castro, 1985]

A Federación de Sociedades Gallegas de Bos Aires, fundada en 1921, axudou economicamente ó PG en diversas ocasións así coma ó Partido Agrario Galego, que lideraba Antón Alonso Ríos. [CORGA: C. Sixirei Paredes, *A emigración*, Galaxia, 1988]

Un lugar solitario e medoñento onde entobaba o lobo, así coma outros animais propios daqueles contornos crebados e montesíos: garduñas, raposos, teixugos, corzos e aves carniceiras, sobre todo corvos, e moreas de serpes. [CORGA: X. Fernández Ferreiro, *Agosto do 36*, Xerais, 1993]

Mais a maioría dos checos, así coma moitos luteranos eslovacos, non aceptaban a "secesión" eslovaca respecto da "nación común" checa. [CORGA: GI1996-07/30]

Porque non se perciben as inxustizas e os agravios comparativos que estes comportamentos provocan, así coma os seus efectos negativos para a democracia. [CORGA: TN1997-11/17]

Igualmente, a cada rapaz, usuario dun comedor escolar, así coma os asistentes ás charlas informativas, se lles entregou un cepillo dental. [CORGA: CG2000-12-24/4]

Sempre dicía que saber ler e escribir, así coma un chisco de contabilidade, era algo necesario para a mellora do campo. [CORGA: A. Riveiro Coello, *As Rulas de Bakunin*, Galaxia, 2001]

O segundo Congreso Internacional de Lingüistas (Xénova 1931), así coma o sexto (París 1948) dedicaron especial atención á Interlingüística. [CORGA: T. Moure, *Outro idioma é posible. Na procura dunha lingua para a humanidade*, Galaxia, 2005]

Todo tipo de obras pasaron pola Chocolataría así coma todo tipo de artistas e persoas dende que en setembro do 2005 se derrubara o muro de chocolate co que o artista Diego Santomé cubriu a porta na inauguración do espazo. [CORGA: *ANT2006-12-14/65*]

Ademais, nunha pregunta dirixida ao Congreso dos Deputados, recollida por Europa Press, o BNG pregúntalle á Administración sobre a súa análise respecto das causas da fermentación do abono, e se o buque estaba "preparado" para transportar este tipo de mercadorías, así coma se a fermentación da carga tivo "algo que ver" coa calor desprendido polas máquinas do barco. [CORGA: *GH2007-03-05/13*]

En segundo lugar, as posibles conmutacións por outros elementos análogos non só nos mostran nexos copulativos (*e*, *e mais*), reforzados ou non coa presenza do adverbio *tamén*:

Neste senso, A Coruña, Ferrol e mais Vigo son as mais problematizadas. [CORGA: *TE1977-34/2*]

Estes e outros aburrimentos sumados, e mais tamén un distante estrondo enciclopedista que lle rebulía no sangue, foi o que levou a Miguel Quintás a esas paixóns científicas que causaron o seu fin. [CORGA: M. A. Murado, *Memoria de derribos*, Galaxia, 1993]

A calma foise restablecendo mercé á intervención do arrufado e veterano mestre Tarfón e a do hazzam e mais tamén a do decurión, os cales, berrando igual có público "Viva Iahvé!", conseguiron ir introducindo frases que convidaban a escoitar. [CORGA: B. Graña, *Protoevanxeo do neto de Herodes*, Xerais, 2006]

senón tamén a posibilidade de introducir conxuncións adversativas reforzadas cun adverbio, que é o que lle confire o valor iterativo:

As fundacións das igrexas fánse entón polos bispos ou frades, pero tamén polos donos das vilas ou polos veciños dunha aldea ou vicus. [CORGA: X. Fariña Jamardo, *A persoalidade da parroquia galega*, SEPT, 1976]

Ao longo dos seis anos que durou a experiencia prantáronse máis de cincocentos frutais, nomeadamente pereiras e maceiras, pero tamén cerdeiras, castiñeiros, nogueiras, mermeleiros, limoeiros, laranxos, etc. [CORGA: N. Gabriel, *Agricultura e escola. Contra a rutina e o éxodo rural*, USC, 1989]

A imaxe dos equipos é un suxeito complicado, no que inflúe a historia, tanto recente como máis lonxana, pero tamén as filias, fobias e os intereses entrelazados con diversos medios de comunicación. [CORGA: *ANT2002-09-19/50*]

xeralmente pesca: fanecas e congros principalmente, mais tamén cabras, saiñeiras e ás veces algunha bertorella. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

e mesmo estruturas comparativas de igualdade como a que conforma o adverbio *igual* seguido da conxunción *ca*, ou ben secuencias semellantes que manifestan comparación:

Nembargantes esta medida non deixa de ser algo que carece dunha verdadeira razón fisiolóxica, pois como é ben sabido no aspecto do peso, igual ca en outros, inflúen unha chea de factores de variación, como a raza, ecoloxía, etc. [CORGA: I. Caamaño, *Unha alternativa gandeira pra os montes da Galiza*, Galaxia, 1981]

Na paz, igual ca na guerra, o home amosa o seu lado terrible. [CORGA: R. Fonte, *As regras do xogo. Dez relatos en xénero negro*, Edicións do Cumio, 1990]

Os outros delegados _Xurxo, entre eles_, igual ca Antonio, eran todos marxistas e activistas políticos de grupos radicais que, como xa dixen, nunca conseguín distinguir. [CORGA: X. Alvilaes, *Memorial de agravios*, Galaxia, 1992]

A incisión, igual ca a perforación espontánea, tende a pecharse soa axiña e non deixa secuelas. [CORGA: C2005-11/11]

O compás, igual ca outros aspectos da organización musical, é de natureza arquitectónica. [CORGA: M. Groba Lorenzo, *Introdución á linguaxe musical*, Galaxia, 2006]

Non obstante, o que acontece é que canto máis raras son as especies, máis valor adquiren, do mesmo xeito ca os selos ou as moedas de coleccionista. [CORGA: C2007-03/25]

Propicio a filias e fobias, tíñalle xenreira a Madrid e á Coruña do mesmo xeito que gostaba de Pontevedra. [CORGA: F. Dopico, *A ilustración e a sociedade galega. A visión de Galicia dos economistas ilustrados*, Galaxia, 1978]

Caso un pouco distinto, pero ao tempo tamén problemático pola súa incerta clasificación entre as estruturas comparativas ou as copulativas, é o *como* da correlación *tanto...como*. Naturalmente, non nos estamos referindo aos casos claros de comparación de igualdade, tipo

Se alomenos gaña tanto como perde, se engade algún aspecto novo, entón si que a fago. [CORGAetq: ANT2005-03-03/49]

A crítica relativa á utilización das axudas non é a única que o Consello fai no seu informe, malia o documento non ser especialmente crítico cunha xestión da que a auditora privada Arthur Andersen expresou "unha opinión con salvidades", algo que na xiria dos auditores é tanto como unha desaprobación. [CORGAetq: TN1998-04/14]

senón a aqueloutros nos que existe unha correlación discontinua co adverbio *tanto*, co que conforma, segundo algúns autores (por exemplo Álvarez e Xove, 2002: 654 ou López Viñas et al., 2010: 399), unha locución conxuntiva copulativa.

Vexamos primeiro algúns exemplos reais de uso extraídos do corpus de adestramento e logo trataremos brevemente a súa caracterización:

Neste sentido, o sistema educativo e de investigación, tanto nas institucións especializadas como nas empresas, terá unha importancia renovada. [CORGAetq: MC1980-12/1]

Aquí parece que tanto Aznar como Felipe González teñen como único referente a doutrina dominante imposta pola democracia cristiá alemana. [CORGAetq: TN1997-07/29]

Por sectores, a evolución en outubro "foi estrictamente idéntica tanto na zona euro como na UE", precisa Eurostat nun comunicado. [CORGAetq: CG2000-12-08/67]

Tanto BNG, como PSOE presentaron as súas queixas publicamente polos escasos recursos que destinará o Goberno Central a Galicia. [CORGAetq: ANT2002-10-03/31]

Con todo, é previsible que o interese por facer o que a un lle gusta ocupe un lugar privilexiado tanto para os pequenos como para os adolescentes: [CORGAetq: GH2008-08-05/29]

O que faría falla é todo o contrario: tanto iniciativas políticas como actuacións institucionais e leis que promovan o idioma. [CORGAetq: GH2008-08-05/43]

Vistos os exemplos, debemos aclarar antes de continuar que o sistema de XIADA non nos permite o tratamento e recoñecemento de locucións discontinuas¹⁸², xa que logo a cuestión a dilucidar é como imos caracterizar o *como* desa correlación.

En primeiro lugar, ningún dos manuais nos que fundamentamos a nosa etiquetaxe (RAG/ILG, 1995, 2003; Álvarez, Regueira e Monteagudo, 1986) recolle entre as locucións conxuntivas copulativas *tanto...como/coma*, nin entre as conxuncións copulativas *como* nin *coma*.

En segundo lugar, é verdade que, a semellanza das conxuncións copulativas,

1) a concordancia esixe que o verbo vaia en plural se os elementos que se correlacionan están en singular e funcionan como suxeito:

¹⁸² Non cuestionamos sequera a posibilidade de se o era ou non, porque aínda que o fose o sistema non nos permitiría tratala como locución.

Por este motivo, os expertos aseguran que a silvicultura debe orientarse cara outras especies porque tanto o eucalipto como o piñeiro pinaster non teñen acabado co principal problema do monte galego, o lume. [CORGAetq: *ANT2005-03-03/56*]

Fronte a isto, tanto o portavoz do PSOE no Congreso, Diego López Garrido, como o coordinador xeral de IU, Gaspar Llamazares, aseguraron que "non pode dicirse que haxa ningún tipo de alerta máxima". [CORGAetq: *GH2007-03-30/51*]

2) se o que se correlacionan son adverbios en *-mente*, por regra xeral, só o segundo deles leva o sufixo:

Máis importante, tanto cualitativa como ideolóxicamente, aparécesenos Reforma Social Española, outro partido-persoal que equivale a nomear a Manuel Cantarero del Castillo, e que é a asociación menos á dereita de todas as autorizadas por Arias Navarro, cunha base inicial de militantes provintes de Falange e Frente de Juventudes. [CORGA: X. I. Taibo e M. Rivas, *Os partidos políticos na Galiza*, La Voz de Galicia, 1977]

Tamén é destacable, neste caso como aspecto negativo, a falla de mobilidade tanto vertical como horizontalmente. [CORGA: A. Teixeira Piñeiro, *O Parlamento de Galicia*, Ir Indo, 1988]

Se actúan en consecuencia, tanto interna como externamente, haberá esa catarse imprescindible para recuperar posicións. [CORGA: *TN1997-11/16*]

Era preciso subir por unha escaleira revirada, empinada e con pasos para cabaleiros de gran zancada, que non era o caso da maioría dos parlamentarios, tanto física como politicamente: pouco alento, curtos de miras. [CORGA: *ANT2006-12-29/1*]

Non obstante, tamén localizamos ocorrencias que transgreden as dúas normas anteriores e que, na nosa opinión, non deixan de ser correctas:

Na súa intervención, Marta Rodríguez recorreu á comparativa con outras autonomías e asegurou que tanto o número de valoracións como o de prestacións adxudicadas chocha co que ocorre noutras comunidades. [CORGAetq: *GH2008-08-05/38*]

A Sociedade acumulou perdas por valor de 2.128 millóns de pesetas, o que a situaba "nun dos supostos de disolución e liquidación" previstos na Lei, como se encarga de advertir tanto o Consello de Contas como a auditoría privada que se inclúe no informe de fiscalización. [CORGAetq: *TN1998-04/12*]

Por iso, a nosa intención é máis ben a de espallar a lexislación correspondente, deixando á interpretación persoal de cadaquén a consideración que lle mereza o problema, mentras non sexa producida a intervención xurídica nun senso estritamente profesional, tanto administrativamente como xudicialmente, "a instancia de parte" por así decilo. [CORGA: X. L. Fontenla Rodríguez e M. Barcia Lago, *Contaminación industrial e desenvolvemento. Análise*

xurídica da situación da Ría de Pontevedra por instalación de industrias contaminantes, SEPT, 1975]

As presentacións son unha vía poderosa para conseguir difundir as nosas mensaxes, tanto verbalmente como visualmente". [CORGA: ANT2008-06-26/21]

Así mesmo, na suposta locución conxuntiva copulativa discontinua, *tanto... como* volve concorrer a variante con *coma* nos mesmos contextos que estamos describindo, en paralelo ou simultaneamente aos de *como*:

Na súa intervención, Santiago Seco, director do Departamento de Análises, Prospeccións e Asesoramento, afondou no significado das ACIS e AXES, tanto para mellora-la calidade do leite coa vista posta no 1º de xaneiro do 98, no caso das primeiras, xa que a partir desa data non se poderá recoller leite que non reuna as condicións mínimas de asepsia segundo dispón a UE, coma para mellora-los resultados económicos das explotacións, coas segundas, e mesmo co servicio de laboratorio que se puxo en marcha. [CORGAetq: F1997-08/2]

No caso de España foi fundamental a labor que levaron a cabo tanto Abelló coma os Laboratorios de Zeltia na súa comercialización. [CORGAetq: ANT2008-07-10/11]

Ela, que aproveitaba todas as representacións mundanas pra amosstrar a súa beleza desexada tanto polos homes coma polas mulleres, non estaba na festa do vello. [CORGA: E. Moreiras, *Fogo solto*, Edicións do Castro, 1976]

"Pois, se quere, ata se lle podería chamar "festival": un festival de traballos para televisión feitos por xente doutras nacionalidades tanto do Estado español coma de Latinoamérica, ou Brasil, ou Portugal... [CORGA: EC1985-EX/7]

Entrevistado: Certamente cando estudiabamos en Madrid, tanto Lorenzo, coma Bernardino, coma min, fomos conscientes de que tiñamos que aproveita-la Escola de Filoloxía Española para cando chegásemos a Galicia funda-la Escola de Filoloxía Galega. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

_Pero tanto ela coma el non gastan moitos fumes. [CORGA: B. Graña, *Protoevanxeo do neto de Herodes*, Xerais, 2006]

As bogas tanto se consumían frescas coma salgadas e tanto no verán como no inverno. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

Ademais, en caso de clasificar como locución conxuntiva copulativa *tanto...como/coma* tamén habería que estender esa análise a *tanto...canto* e amais a *non só...senón tamén*, pois a significación e funcionalidade parecen idénticas:

Entón, eu non sei cómo será o seu porvir, mais o que sei é que ten funcionado positivamente tanto na mostra do ano pasado en Baiona canto nesta de aquí, en Vigo. [CORGA: *MC1981-03/2*]

De todas formas, parece atinada tanto a bipartición do tempo canto o feito de situa-lo comenzo do ano agrario en outubro, tralos traballos de sega (xullo), malla (agosto) e vendima (setembro). [CORGA: X. Chao Rego, *Para comprendermos Galicia*, Galaxia, 1988]

Os interrogantes que se abren refírense a se estará o nacionalismo, nesta lexislatura que se acaba de abrir, en condicións para marcar o roteiro para o cambio político en Galiza, tanto na dirección dun cambio para a esquerda canto na dirección de dar un salto adiante a partir de estarmos no goberno autónomo, para avanzar na consecución de institucións políticas realmente capaces de exercer o autogoberno. [CORGA: *TN2005-11/14*]

Mais moito temo que nisto, como noutras tantas cousas, percíbese doadamente a excesiva influencia dos teólogos que, ao veren minorar o peso dos seus discursos fronte ao poder que as novas ciencias van gañando, tanto nas mentes ilustradas canto entre os máis simples, queren perfilar e buscar explicacións morais para feitos que só materialmente poden probarse. [CORGA: T. Moure, *Herba Moura*, Xerais, 2007]

"O PCG faise intérprete e defensor dos intereses e arelas da maioría absoluta da nosa sociedade, non só polo que respecta ó presente momento histórico, senón tamén polo que se refire ó porvenir. [CORGA: X. I. Taibo e M. Rivas, *Os partidos políticos na Galiza*, La Voz de Galicia, 1977]

As consecuencias traumáticas desta situación fanse notar non só no aspecto social senón tamén no psicolóxico e mesmo siquiátrico. [CORGA: *EC1985-17/11*]

O mundo do Seminario era, como dixen, rural non só pola procedencia de todos nós, senón tamén pola fonte de información do mundo e pola escasa curiosidade intelectual. [CORGA: X. Alvilares, *Memorial de agravios*, Galaxia, 1992]

Eu conducía non só o coche senón tamén a conversa e, como imaxinan, falabamos sobre a bicicleta. [CORGA: *CG2000-11-01/37*]

Pero a película non só está dirixida ao público infantil senón tamén ao adulto. [CORGA: *GH2008-08-06/50*]

Algo que pode botar un pouco de luz sobre a categorización do *como* ou *coma* como segundo termo da correlación que estamos describindo é ver se as restricións que se aplican ás conxuncións comparativas (*coma* máis pronome persoal tónico e *como* seguido de verbo) se manteñen *grosso* modo nestes contextos.

No corpus de adestramento non se dá ningún caso desta estrutura cuxo segundo termo sexa un pronome e si dúas ocorrencias nas que o que se une son dúas cláusulas, aparecendo en ambas as dúas *como*:

O novo Estatuto da Viña e do Viño tanto pode seguir coas denominacións de orixe reformadas como pode arrombalas. [CORGAetq: CG2001-10-08/1]

Por eso mesmo, constitue un momento de inflexión que tanto pode ter como consecuencia a reprodución dos actuais centros públicos e privados de poder e decisión económica e política _dando lugar a unha nova acumulación, baixo formas renovadas de dominio tecnolóxico, comercial e financeiro_ como servir para propiciar o inicio dunha nova fase de desenvolvemento, fora da lóxica da dependencia político-económica e non baseada simplemente nas necesidades da acumulación de capital, senón nas necesidades sociais, na difusión dos coñecimentos científicos, tecnolóxicos e culturais, e na liberdade dos pobos. [CORGAetq: MC1980-12/1]

Na consulta ao CORGA, debido á imposibilidade de utilizar etiqueta ou lema coas dificultades que isto provoca nunha pescuda coma a que pretendemos¹⁸³, decidimos prescindir dos verbos como termos da estrutura *tanto...como/coma* e limitar a busca ao pronome de primeira persoa singular e ao de terceira, singular ou plural, entendendo que os resultados serán o suficientemente representativos para demostrar a exclusión de *como* no segundo termo da estrutura bimembre e a esixencia de *coma* cando o termo do segundo elemento da estrutura é un pronome persoal tónico. A nosa pretensión, como se pode deducir, é expoñer a proximidade desta estrutura coa comparativa e xustificar así a elección da etiqueta Cs (*Conxunción subordinante*) para *como* e *coma* que empregamos na etiquetaxe.

Os resultados¹⁸⁴ demostránnos que o emprego de *como* seguido de pronome persoal tónico é insólito, pois os únicos casos que atopamos en case 26 millóns de formas son os 5 seguintes, ocorrencias ademais que, como falantes, rexeitaríamos:

¹⁸³ A falta de etiquetaxe no CORGA provoca que a consulta resultante, independentemente de como se formule, proporcione todos os casos de *como*: relativos, conxuntivos, interrogativos-exclamativos, presentes en locucións, etc., así como que non poidamos buscar *como/coma* seguido de forma verbal. A consecuencia máis inmediata disto é que teríamos que revisar todos os casos manualmente, un por un, sen outra posibilidade de peneirar os resultados obtidos.

¹⁸⁴ A posibilidade de utilizar operadores booleanos permitiunos obter xa todas as ocorrencias que cumprisen co criterio de apareceren nelas as formas *tanto* e *como* seguido de *eu* –en consultas sucesivas, min, el*– que estivesen a unha distancia de quince palabras, como moito. A primeira consulta realizada, en concreto, foi: NEAR ((tanto,como eu), 15)

Nun dicir amén, tanto el como eu eramos outros, dúas persoas cheas de harmonía. [CORGA: X. Fuentes Castro, *O pescador de cana*, Toxosoutos, 2006]

A Morte: tanto el como elas están feitos para a danza. [CORGA: C. Freire Pérez, "*S.O.S. mar de fondo*", Fontel, 1994]

Ante a insistencia dos medios, Cristina Almeida amosouse realmente enfadada pedindo que os deixasen pasar porque "é un dereito mínimo" e advertindo que se tanto o seu cliente como ela recibían algún empurrón máis, denunciaría os fotógrafos, cámaras e demais xornalistas. [CORGA: *CG1996-02-07/159*]

Claro, despois disto é loxico que saia Serra dicindo que para esta viaxe non facían falta alforxas, e que el xa dicía que era un incomprendido, e que tanto García Vargas como el mesmo non eran delincentes senón heroes. [CORGA: *CG1996-02-10/6*]

Pola súa parte, o ministro de Facenda, Cristóbal Montoro, asegurou onte que tanto Rato como el responderán a tódalas preguntas que lles sexan formuladas durante a súa comparecencia o próximo luns ante a comisión. [CORGA: *CG2001-10-26/35*]

Pola contra, danse 38 ocorrencias con *coma*: 14 co pronome de terceira e 24 co de primeira singular. Neste último caso aparece ademais a variación na expresión entre a forma recta, 3 casos, e a oblicua, 21. Vexamos un par de exemplos de cada:

Porque en verdade acontecía que, sen térense posto de acordo, tanto a nai coma el mantivéronse na mesma actitude, teimando por cadansúa banda no agachamento, en refuxiárense decote nunha amnesia absoluta... [CORGA: M. Riveiro Loureiro, *O Patrón*, Edicións do Cumio, 1990]

Pero tanto ela coma el non gastan moitos fumes. [CORGA: B. Graña, *Protoevanxeo do neto de Herodes*, Xerais, 2006]

Tanto o Amadeo coma eu, tiñamos ganas de leria aquela noite. [CORGA: X. Rodríguez Barrio, *A fundición e outras narracións*, Edicións do Castro, 1978]

Sen embargo, tanto el coma eu sabiamos que era esa a única saída, a únida decisión que tiña sentido, e, máis tarde ou máis cedo, íamola tomar... [CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

Tanto o Atravesado coma min, xa non podiamos facer máis do que é tratarmos de saír, se podiamos, daquel labirinto. [CORGA: S. Álvarez, *Memoria da guerrilla*, Xerais, 1991]

Pero tanto o teu avó coma min, tiñamos familia alá. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

Curiosamente, o que algúns autores consideran estrutura bimembre fixa (RAE, 2009: 2411) cuxo segundo termo non pode ser repetido, aínda que si permite a coordinación no seu interior, aparece refutado por dous exemplos no CORGA:

Entrevistado: Certamente cando estudiabamos en Madrid, tanto Lorenzo, coma Bernardino, coma min, fomos conscientes de que tiñamos que aproveita-la Escola de Filoloxía Española para cando chegásemos a Galicia funda-la Escola de Filoloxía Galega. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

Xoán López Facal: Quizais influíse que tanto Xoaquín, meu irmán, coma min, coma Camilo, eramos enxeñeiros industriais. [CORGA: S. de Toro, *Camilo Nogueira e outras voces*, Xerais, 1991]

É obvio, polo que acabamos de ver, que tanto *así como/coma* canto *tanto...como/coma* teñen valor coordinante, mais tamén parece clara a súa relación coas estruturas comparativas. Nós sostemos a hipótese de que a partir da comparación de igualdade que subxace nelas se propaga o significado copulativo. Por suposto, sería necesaria unha análise moito máis dilatada da cuestión para extraer conclusións en firme; non obstante, a nosa intención, repetimos unha vez máis, non é realizar un estudo profundo sobre o *como*, senón só intentar xustificar a etiquetaxe que practicamos e proporcionar material para que logo os investigadores, ao recuperar a información, analicen, decidan e expliquen os criterios de significación e diferenciación desta unidade nas súas diferentes realizacións.

Finalmente, para rematar con esta epígrafe, só nos queda comentar a lematización do adverbio interrogativo-exclamativo e destacar o criterio de diferenciación establecido entre o *como* adverbio interrogativo-exclamativo nas interrogativas indirectas e o *como* conxunción introdutor de cláusulas subordinadas de complemento directo.

No que se refire ao primeiro punto, para os interrogativos-exclamativos indirectos, xa se tratou nos capítulos precedentes o cambio acentual que introduce a última edición das Normas (RAG/ILG, 2003: 31-32): pasou de ser obrigatorio a recomendarse o seu uso só en casos de ambigüidade. Do que non falamos naquel momento, por non ser oportuno, foi da súa lematización, aspecto que trataremos a continuación e que malia exemplificar con *como* é aplicable a todos os interrogativos-exclamativos.

Dado que no CORGA hai documentos anteriores á primeira edición das *Normas* (RAG/ILG, 1982) e que non todos os autores respectan a normativa vixente, atopámonos xa antes do cambio establecido no 2003 coas formas ortográficas *como* e *cómo*, tanto nas interrogativas-exclamativas directas como nas indirectas:

O concepto da economía no inmediato futuro non se vai centrar tanto en crecer, senón en como crecer. [CORGAetq: *GI1996-06/30*]

¿Como é posíbel acadar unha rendibilidade tan alta? [CORGAetq: *ANT2002-10-10/33*]

¿Cómo constituír procesos e institucións colectivas para facer valer na vida social opcións que teñen profundos arraigamentos en moitas persoas? [CORGAetq: *TN1997-06/7*]

Catro anos despois, sentado fronte a el, pregunteille cómo fora. [CORGAetq: *CG2002-05-15/17*]

Para a súa lematización consideramos que son variantes en distribución complementaria, e decidimos remitir as dúas formas ao lema *como*, o que se emprega nas directas, a semellanza por exemplo dos alomorfos do pronome de terceira acusativo que se remiten a *o*, a considerada primeira forma. Así, a etiquetaxe das expresións anteriores dos catro *como* redúcense no noso sistema a dúas análises distintas, consecuencia da diferenza na acentuación que se mantén na forma, mais remítense a un só lema:

como	Wg	como
cómo	Wg	como

O *como* presente nas interrogativas-exclamativas directas non presenta ningún problema para a súa identificación e posterior caracterización, motivo polo que o obviamos nesta descrición. Agora ben, cuestión moi distinta, na que tamén inflúe decisivamente o cambio operado na normativa de 2003, é a distinción entre o *como* completivo e o *como* interrogativo indirecto.

En teoría, ata a publicación da última edición das *Normas* (RAG/ILG, 2003), se no interior dunha secuencia aparecía un *cómo* acentuado, este tiña que ser necesariamente adverbio interrogativo, pois a normativa prescribía a súa acentuación para evitar anfiboloxías. Así, etiquetamos como Wg (*Adverbio interrogativo-exclamativo*) os seguintes casos:

No segundo caso, se quere que España estea entre os primeiros da moeda única, o PP debe explicar" cómo vai logra-las reducións do déficit necesarias para cumprir cos criterios e o descenso dos impostos", ó mesmo tempo. [CORGAetq: CG1996-02-14/109]

A pregunta que se pode formular é cómo se emendaron tan contiosas perdas. [CORGAetq: TN1998-04/12]

Ademais, xa non hai que incluír xustificante ningún dentro do sobre da declaración, salvo en Navarra e no País Basco, precisa a Guía, que explica a través de 255 cuestións cómo facer correctamente a declaración e, á vez, cómo pagar menos impostos dentro dos límites legais. [CORGAetq: CG1999-05-07/25]

Non obstante, como xa vimos na primeira parte deste capítulo, nos textos do CORGA non sempre se respecta a norma, ben por convicción ben por omisión involuntaria. O caso é que tamén se dan ocorrencias de *cómo* que, na nosa opinión, non son interrogativas. Por exemplo:

Despois de ver cómo os portugueses se introducen en Galicia con forza (Cimpor compra Corporación Noroeste; Caixa Xeral de Depósitos adquire o Simeón), un alédase ó comprobar que tamén as empresas galegas se expanden (Caixa Galicia compra parte do Urquijo; San José adquire parte de Huarte). [CORGAetq: CG1996-02-12/5]

O banco viu cómo o seu beneficio neto no primeiro trimestre se reducía en máis dun 78% con respecto ós 481,8 millóns de euros conseguidos nese período do 2001. [CORGAetq: CG2002-05-20/9]

E isto introdúcenos de cheo na diferenciación entre o *cómo* interrogativo indirecto e o *como* introdutor de subordinadas substantivas de C.D. Se antes do 2003 podíamos pensar inxenuamente que o criterio da acentuación nos axudaría a discriminar ambos os valores, criterio que xa invalidan os dous exemplos anteriores, despois da publicación da última edición das *Normas* (RAG/ILG, 2003) a diferenza que puidera establecer a regra suprasegmental desaparece.

Sabemos que, fronte á tonicidade do *como* interrogativo e da súa permuta por *de que maneira*, o *como* introdutor de C.D. é átono, emprégase con verbos de percepción, sobre todo *ver* e *oír*, e é conmutable pola conxunción *que*. Velaí algúns exemplos:

"...sentes como os ollos se che van convertendo en mar aos poucos. [CORGAetq: ANT2005-03-03/70]

Eu vexo como moitos rapaces sofren na clase cando se empregan encerados electrónicos. [CORGAetq: ANT2008-07-10/50]

Non nos queremos dar conta do sufrimento que tiveron moitas xeracións anteriores que vían como perdían os fillos, como desaparecían familias enteiras. [CORGAetq: ANT2008-07-10/11]

En total, segundo informan as ONG que traballan na rexión, calcúlase que son preto de 4,5 millóns as persoas afectadas polo conflito que viron como se destruían os seus medios de vida tradicionais. [CORGAetq: GH2008-08-05/97]

Con todo, a noción de modo, aínda que secundaria, segue presente, polo que ás veces, por interferencia co adverbio interrogativo, se acentúa erroneamente. Esta é a hipótese que barallamos para soste que, malia aparecer a forma acentuada, os dous casos de “ver + cómo” de arriba se corresponden co *como* conxunción subordinante introdutor de subordinadas substantivas de complemento directo, polo que as etiquetamos no corpus de adestramento como Cs.

O anterior non implica que todas as ocorrencias de “ver + como” ou “ver + cómo” reciban a etiqueta de conxunción subordinante, xa que nin a acentuación nin a aparición con verbos de percepción son criterios válidos para a identificación do *como*; o principio sobre o que sostemos o recoñecemento desta forma tan incómoda é a súa conmutación por *que* ou por *de que maneira*. Así, nas secuencias que seguen, o primeiro recibe a etiqueta Wg (*Adverbio interrogativo-exclamativo*), mentres que o segundo acolle a de Cs (*Conxunción subordinante*):

ver + Wg

Os responsables económicos da Eurozona comprometéronse a vixila-la evolución deste problema unha vez ó ano para ver cómo o trata cada país, dado que na mellor das hipóteses o gasto público en pensións aumentará unha media de entre o 4 ou o 5 por cento do Producto Interior Bruto (PIB). [CORGAetq: CG2000-11-08/71]

ver + Cs

Ao igual que no caso de Sheryl Crow, a artista viu como este proxecto se conxelou cando lle diagnosticaron un cancro. [CORGAetq: DLV2007-05-15/37]

4.4.4. **Que**

Acabamos de ver no apartado anterior que *que* é a segunda forma máis frecuente da versión 1.5 do CORGA, con 979 763 ocorrencias, só superada polo *token de*, e intuimos que esa multipresenza ten o seu correlato no CORGAetq, malia que xa desagregada en función da categoría gramatical. E así é: a segunda posición do corpus sen etiquetar, onde só vemos as formas sen posibilidade de acceder ás etiquetas que as

caracterizan, transfórmase na tripla división do *que*, segundo a súa inclusión en clases diferentes, ocupando o posto oito, nove e douscentos trinta da listaxe de elementos gramaticais da versión 2.4 do CORGAetq ou corpus de adestramento de XIADA:

Orde	Elemento gramatical	Categoría	Frecuencia	Porcentaxe	Porcentaxe acumulada
1	de	Preposición	32969	7.7383	7.7383
2	o	Artigo	22809	5.3536	13.0919
3	a	Artigo	17527	4.1138	17.2057
4	en	Preposición	12177	2.8581	20.0638
5	e	Conxunción	8972	2.1059	22.1696
6	os	Artigo	7608	1.7857	23.9553
7	a	Preposición	7205	1.6911	25.6465
8	que	Relativo	6978	1.6378	27.2843
9	que	Conxunción	5224	1.2261	28.5104
230	que	Interrogativo/Exclamativo	130	0.0305	50.8887

Naturalmente, a maiores da desagregación provocada pola adscrición a unha categoría gramatical, existe unha diferenza lóxica no cómputo das ocorrencias de *que* entre o CORGA e o CORGAetq: mentres que no primeiro calquera aparición de *que* entre dous espazos en branco vai sumar no total, no segundo só contabilizarán aquelas cuxo lema sexa *que*. É dicir, no CORGA computan os *que* de *aínda que*, *posto que*, *xa que*, *mentres que*, etc., mais no CORGA etiquetado non sucede así, estas unidades conformarán unha locución conxuntiva independente, propia:

Orde	Elemento gramatical	Categoría	Frecuencia	Porcentaxe	Porcentaxe acumulada
123	aínda que	Locución	210	0.0493	46.8169
144	mentres que	Locución	192	0.0451	47.7983
150	xa que	Locución	184	0.0432	48.0635
...					

Sen dúbida, os culpables directos da complexa tarefa de desambiguación que supón calquera ocorrencia de *que* non o son tanto os diferentes valores da forma en si coma nós mesmos. É verdade que en numerosas ocasións resulta, cando menos,

problemática a súa clasificación nunha das tres categorías gramaticais á que o remitimos: Conxunción, Interrogativo-exclamativo e Relativo, mais se prescindiramos da pertinencia dos atributos *valor*, *xénero* e *número* moitos dos problemas de etiquetaxe evitaríanse.

Para que nos fagamos unha idea da complicación que indicamos, imos recoller de seguido todas as etiquetas que en XIADA, en teoría, é posible asignarlle á forma *que*:

Cs	[Conxunción, subordinante]
Cc	[Conxunción, coordinante]
Gdaa	[Interrogativo-exclamativo, determinante, masculino/feminino, singular/plural]
Gdap	[Interrogativo-exclamativo, determinante, masculino/feminino, plural]
Gdas	[Interrogativo-exclamativo, determinante, masculino/feminino, singular]
Gdfp	[Interrogativo-exclamativo, determinante, feminino, plural]
Gdfs	[Interrogativo-exclamativo, determinante, feminino, singular]
Gdmp	[Interrogativo-exclamativo, determinante, masculino, plural]
Gdms	[Interrogativo-exclamativo, determinante, masculino, singular]
Gnaa	[Interrogativo-exclamativo, non determinante, masculino/feminino, singular/plural]
Gnap	[Interrogativo-exclamativo, non determinante, masculino/feminino, plural]
Gnas	[Interrogativo-exclamativo, non determinante, masculino/feminino, singular]
Gnfp	[Interrogativo-exclamativo, non determinante, feminino, plural]
Gnfs	[Interrogativo-exclamativo, non determinante, feminino, singular]
Gnmp	[Interrogativo-exclamativo, non determinante, masculino, plural]
Gnms	[Interrogativo-exclamativo, non determinante, masculino, singular]
Tnaa	[Relativo, non determinante, masculino/feminino, singular/plural]
Tnap	[Relativo, non determinante, masculino/feminino, plural]
Tnas	[Relativo, non determinante, masculino/feminino, singular]
Tnfp	[Relativo, non determinante, feminino, plural]
Tnfs	[Relativo, non determinante, feminino, singular]
Tnmp	[Relativo, non determinante, masculino, plural]
Tnms	[Relativo, non determinante, masculino, singular]

Se as contamos, veremos que son 23 etiquetas distintas, pois o feito de variar nelas algún dos atributos que determinan a súa caracterización convérteas para o sistema en 23 posibilidades diferentes. Por suposto, é cuestionable por que se decide que os atributos *valor*, *xénero* e *número* son pertinentes na clase dos relativos e na dos interrogativos-exclamativos. Non obstante, dende o punto e hora en que algúns dos membros destas dúas clases responden a eles formalmente por si mesmos, como demostra a flexión de xénero e número (*canto, canta, cantos, cantas; cal, cales...*) ou a súa realización na lingua (*Canta auga!, Cantos queres?...*), e tendo en conta que tamén nos substantivos ou adxectivos que presentan sincretismo no xénero ou no número, por poñer só dous exemplos incuestionables, si son pertinentes e a ambigüidade se desfai grazas á información que nos prestan os elementos que os rodean, por que non ser coherentes e aplicar tamén o mesmo criterio na categoría Relativo e Interrogativo-exclamativo? Acaso imos prescindir deles polo único feito de que na tradición gramatical da que partimos eses valores non son valedeiros en todos os membros da clase? Imos deixar no camiño a posibilidade de estudar a concordancia entre o pronome relativo e o seu antecedente só porque *que*, illado, é invariable? Imos empecer o estudo dos distintos tipos de palabras que forman unha unidade co interrogativo-exclamativo? Que é demasiada información para algún tipo concreto de investigación? Ten fácil solución: ignórense os valores referidos aos atributos *valor*, *xénero* e *número* e préstese só atención á clase e ao tipo. A acción contraria sería altamente custosa, dende logo imposible de realizar de xeito automático, e manualmente obrigaríanos a volver sobre cada un dos casos. A decisión tomada parece, xa que logo, congruente cos obxectivos e principios polos que se rexe XIADA: facilitar a recuperación de información morfolóxica e léxica do galego actual habilitando un sistema que permita a escalaridade na consulta e na obtención de datos.

Das 23 posibilidades teóricas existentes para o lema *que*, no corpus de adestramento, polo de agora, documéntanse 16 para o token *que* e outra máis só recollida para a variante gráfica interrogativa-exclamativa *qué*:

que	Cc	que	que	Gdmp	que
que	Cs	que	que	Gdms	que
que	Gdfp	que	que	Gnaa	que
que	Gdfs	que	que	Gnas	que

que	Gnms	que	que	Tnfs	que
que	Tnaa	que	que	Tnmp	que
que	Tnap	que	que	Tnms	que
que	Tnas	que	qué	Gdaa	que
que	Tnfp	que			

As que non se manifestaron no subcorpus extraído para a prensa, pero si contamos con que o fagan na ficción, son:

Gdap	[Interrogativo-exclamativo, determinante, masculino/feminino, plural]
Gdas	[Interrogativo-exclamativo, determinante, masculino/feminino, singular]
Gnap	[Interrogativo-exclamativo, non determinante, masculino/feminino, plural]
Gnfp	[Interrogativo-exclamativo, non determinante, feminino, plural]
Gnfs	[Interrogativo-exclamativo, non determinante, feminino, singular]
Gnmp	[Interrogativo-exclamativo, non determinante, masculino, plural]

Formalmente non podemos establecer ningunha regra lingüística que facilite a identificación e caracterización do *que*, xa que coa excepción da variante *qué*, que fóra erros vai ser sempre exclamativa-interrogativa, a unidade que antecede a este *token* non permite a súa adscripción xenérica a unha categoría dada. Así é que, tanto para a conxunción subordinante como para o relativo, atopámonos con que os contextos nos que ambas as etiquetas son posibles son moi semellantes, polo que, como é lóxico, a porcentaxe de acerto na desambiguación automática é baixa¹⁸⁵.

Que os contextos son bastante similares demostrano os seguintes grupos de secuencias. Todos os exemplos de a) corresponden á conxunción subordinante, mentres que nos de b) trátase sempre dun relativo:

1. Substantivo + que:

a)

"Desde logo eu xa o dixen, e díxoo o ministro de Facenda, que o Goberno seguirá con grande atención que as reducións do custo do cru repercutan á máxima velocidade en España", dixo.
[CORGAetq: CG2000-12-21/43]

¹⁸⁵ A falta dunha análise específica, valémonos das nosas apreciacións como revisores manuais da etiquetaxe para este aserto. A porcentaxe de acerto tendo en conta a etiqueta completa, e non só a clasificación categorial, calculamos, estará entre un 50-50 ou un 60-40.

Eu sempre aconsello aos alumnos que tendo a carreira feita en calquera conservatorio vaian a outros países a perfeccionarse. [CORGAetq: ANT2005-03-03/49]

Din os analistas que o camiño do bilingüismo efectivo faise máis curto día a día. [CORGAetq: GH2007-03-17/84]

De feito, a lapela do libro infórmanos de que A casa de Mango Street se utiliza en moitas escolas dos Estados Unidos como lectura obrigatoria, o cal, dados os tempos que corren, soa máis coma unha ameaza que coma un eloxio. [CORGAetq: ANT2008-07-10/30]

b)

Actitude que só se concibe en sociedades moi desenvolvidas, formadas e conscientes. [CORGAetq: GI1996-06/30]

O pleno do Parlamento Europeo aprobou onte os acordos que alcanzaron os Quince en Berlín sobre a Axenda 2000, o marco financeiro da Unión ata o ano 2006. [CORGAetq: CG1999-05-07/15]

Amenábar é ademais artista porque foi quen de recrealo nunha moi boa película, rodeado de actores que souberon entendela. [CORGAetq: ANT2005-03-03/50]

O accidente que orixinou a vertedura produciuse a primeira hora do luns 28 de xullo no lugar de Trapedra cando o condutor dun camión da empresa lucense Manfor perdeu o control do seu vehículo, que quedou envorcado sobre o seu lado dereito. [CORGAetq: GH2008-08-05/16]

2. Adxectivo + que

a)

é un obxectivo adicional que se reduzan as diferencias interrexionais nos niveis de desenvolvemento, medidas a través da renda per capita. [CORGAetq: GI1996-05/55]

Mentres, o conselleiro delegado do BBVA, José Ignacio Goirigolzarri, reiterou onte nunha reunión con analistas financeiros que as contas secretas do banco en paraísos fiscais nunca serviron para facer pagos ilícitos, ó tempo que negou calquera irregularidade na compra das súas filiais americanas. [CORGAetq: CG2002-05-09/5]

O portavoz do BNG falou de "extrema dereita" e o líder da oposición estivo máis amornado que nas súas intervencións no Parlamento. [CORGAetq: DLV2007-05-15/18]

Pensamos que a liña é asumir en círculos concéntricos que temos unha lingua útil e viva da que sentirmos orgullosos e que só ten futuro se ten falantes. [CORGAetq: ANT2008-07-10/15]

b)

De nada serviu a porta aberta que a autoridade monetaria estadounidense deixou a unha posible rebaixa de tipos en xaneiro e, logo de comprobar que os índices neiorquinos abrían onte á

baixa, os mercados do Vello Continente caeron sen excepción. [CORGAetq: *CG2000-12-21/46*]

Estou a falar da realidade sobre o sector acuícola que, cando parece que todo está medianamente solucionado ou encarrilado, vai alguén e ¡chas!, as cousas deben volver cambiar o seu rumbo. [CORGAetq: *GH2007-03-30/6*]

Os cartóns amarelos que viron o pasado domingo en Balaídos Gustavo López e Baiano impediranlles xogar no decisivo e definitivo partido que os celestes teñen o próximo domingo en San Sebastián. [CORGAetq: *DLV2007-05-15/103*]

Os índices máis actuais que manexan que din da situación do galego? [CORGAetq: *ANT2008-07-10/15*]

3. Verbo + que

a)

En Galicia pódese considerar que eses cambios afectaron ó sistema financeiro de forma simultánea, aínda que é razoable pensar que se manifestan nas áreas urbanas e entre os clientes con maior capacidade financeira dun xeito máis inmediato. [CORGAetq: *GII1996-06/40*]

Cada vez é máis fácil ver que os liberais _e as forzas económicas_ veñen aproveitando as reformas ligadas ao proceso de integración para impoñer o seu modelo económico e social, pero que é posible pensar outro ou outros modelos diferentes. [CORGAetq: *TN1997-07/29*]

Nunha entrevista concedida ó xornal Avui, Rato aclarou que "a capacidade normativa" non é posible traspasala "pero a recadación si". [CORGAetq: *CG2000-11-20/44*]

Acho que a novela contemporánea é a sublimación da tradición. [CORGAetq: *ANT2008-07-10/27*]

b)

Os contribuíntes con obrigaón de pagar que opten por apraza-la liquidación deberán aboa-lo 60 por cento da cantidade correspondente antes do 21 de xuño e o 40 por cento restante o 5 de novembro como moi tarde. [CORGAetq: *CG1999-05-03/39*]

Unha vez que lle tocara a esta de Colmeiro (que lla mercara a María Antonia, na Galería Torques; si, si, lembrábase ben), notou casualmente certa sensación difícil de explicar que non deixou de chafucalo, polo que, como empurrado por algo, seguiu a observar ata que comprobou debidamente que a engulemia lle desaparecía case ó instante en presenza da pintura. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Non escoitan ó pai de familia que berra dende a cama "Que carallo andas a facer, Jorgito", nin tampouco a voz dun neno asustado e a piques de chorar que lle contesta "Eu nada, papá, é que

Merchi xa me sacou a bolsa da cama". [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Por veces, na diana e alborada que logo demos, percorrendo toda a vila, aínda nos custaba o noso reprimir as ganas de rir que nos viñan ó mirarlle para os beizos, que de cando en cando apertaba, caso de ter que intensificar os redobres. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

4. Adverbio + que

a)

Anunciaron ademais que a próxima reunión do organismo se realizará a finais deste mes e que nela tratarán de aproba-las contas e a memoria do ano pasado. [CORGAetq: *CG1996-02-08/53*]

Lese por aí que as Caixas non apoian como deberan a renovación e modernización do tecido productivo de Galicia. [CORGAetq: *GI1996-06/30*]

O presidente de Unesa advertiu tamén que "o regulador ha de ter sumo coidado e flexibilidade á hora de fixa-las tarifas para non distorsiona-la eficiencia na asignación dos recursos". [CORGAetq: *CG2001-10-10/30*]

Falta agora que os idosos decidan ler, eles tamén, a cara descuberta. [CORGAetq: *ANT2008-07-10/28*]

b)

E saben que despois de pescaren eles, virán outros detrás que tamén necesitarán da pesca para sobrevivir". [CORGAetq: *TN1997-04/7*]

Agora que o reloxo da vida de Reinoso se detivo para sempre pode ofrecer certo interese reproducir algunhas das ideas daquela conversa co, había pouco, nomeado conselleiro-delegado no seu despacho da Coruña. [CORGAetq: *CG2002-05-15/17*]

A tradición foi atacada moitas veces porque confundíase tradición co peor da tradición pero no pouso de séculos da nosa cultura hai máis que aproveitar que de tirar para fóra. [CORGAetq: *ANT2008-07-10/27*]

Pero hainos tamén que aproveitarían o seu tempo libre da xubilación para estudar ou cociñar, ou incluso para facer punto de cruz. [CORGAetq: *GH2008-08-05/30*]

5. Preposición + que

a)

O consello regulador tamén informou de que dúas prestixiosas revistas do sector de viños, The wine advocate e International Wine cellar cualifican como excelentes e sobresaíntes os 14 albariños acollidos á denominación de orixe Rías Baixas. [CORGAetq: CG2000-12-28/12]

Para Ceferino Díaz esta connivencia entre ámbolos dous poderes leva a que "practiquen o principio insolidario e chulesco do eu poño e ti pagas, xa que hai numerosas obras que terán que ser co-financiadas pola Xunta". [CORGAetq: ANT2002-10-03/31]

Logo de once anos traballando a un ritmo frenético, o ferrolán asegura que continuará así "até que aguante". [CORGAetq: GH2007-03-30/40]

Insisto en que é un sen sentido. [CORGAetq: ANT2008-07-10/15]

b)

Cara ás 14.45 horas GMT, a moeda única cambiábase en Frankfurt a 0,8664 dólares, comparado cos 0,8731 dólares a que chegou trala acción. [CORGAetq: CG2000-11-07/78]

Porén, hai ocasións en que o máximo responsábel da factoría que máis diñeiro move de Galicia deixa que a súa voz se escoite fóra do despacho do polígono de Balaídos e utiliza os xornais para enviar determinadas mensaxes. [CORGAetq: ANT2005-03-03/17]

O xogador serbio saltou ó Estadio de Cayo Biscaíno moi concentrado en non repetir os erros que lle levaron a perder en Indian Wells e na segunda ocasión de que dispuxo logrou romper o servizo do xogador español para colocarse cun 3-1 ó seu favor. [CORGAetq: GH2007-03-30/75]

O alto custe da súa renovación, unido as dificultades con que se encontran os poucos canteiros que hai, fai que a meirande parte dos mesmos rematen sucumbindo ante o abandono ao que se ven abocados. [CORGAetq: ANT2008-07-10/65]

6. Conxunción + que

a)

O industrial afirmou que se deron pasos firmes, pero que seguen sen concretarse as ofertas. [CORGAetq: CG1996-02-24/45]

Isto supón que Caixa Galicia contará co 2,5% de Ebro Puleva trala fusión e que terá un conselleiro na cúpula do novo grupo. [CORGAetq: CG2000-11-08/10]

As mesmas fontes afirmaron que a equivocación será corrixida nuns días, pero que descoñecían se os contribuíntes que recibiron información errónea recibirán unha nova comunicación, cos datos verdadeiros, nas próximas semanas. [CORGAetq: CG2002-05-14/14]

A) Éxito das mensaxes propagandísticas que fan creer á sociedade que paga máis que antes e/ou que a iniciativa privada ha de superar á provisión de servizos públicos. [CORGAetq: ANT2002-11-07/29]

b)

Estas son algunhas das conclusións do informe sobre a rebaixa do IRPF, que o ministro de Facenda, Cristóbal Montoro presentou antonte ante o Consello de Ministros e que onte lle explicou á opinión pública. [CORGAetq: CG2000-11-26/38]

"De momento só contamos co apoio do señor [Camilo] Nogueira, que nos facilitou contactar con OPMEGA e que está a axudarnos moito", explican, aínda que reclaman máis axudas "porque en definitiva isto supón exportar un produto e un nome, o Moule de Galice, que tamén beneficia os galegos". [CORGAetq: ANT2002-10-31/25]

Falamos de algo moi particularmente xaponés ou que xa se estendeu polo mundo? [CORGAetq: ANT2005-03-03/52]

Agradar a todos é imposible polo que hai que saber como facer cousas que son necesarias pero que non gustan. [CORGAetq: DLV2007-05-15/20]

Os criterios de desambiguación manexados para o recoñecemento do *que* relativo son os seguintes:

- son elementos anafóricos que se caracterizan por posuír un antecedente explícito, co cal poden concordar formalmente en xénero e número;
- desempeñan unha función sintáctica dentro da cláusula que introducen;
- poden conmutar por *o cal, a cal, os cales, as cales*.

Nas caracterizacións que realizamos asignámoslle ao *que* o xénero e número do seu antecedente, estean estes valores explícitos por medio do artigo ou non. Así, nos exemplos seguintes, recolleemos unha mostra das diferentes variacións coas que nos atopamos no CORGAetq:

Tnfs [Relativo, non determinante, feminino, singular]:

Noutros casos, aparece presupostado unha obra que non ten prazo aínda de execución. [CORGAetq: ANT2002-10-03/31]

Por actividades, a construción é a que rexistra máis accidentes e os seus traballadores, con o 34% do total, sofren o 57 por cento dos accidentes. [CORGAetq: CG2000-11-16/6]

Contraponse, sen embargo, con outra imaxe ben distinta, esa que a veu definindo nos últimos anos como un país que posúe un gran potencial de innovación e creatividade. [CORGAetq: *TN1997-05/1*]

Tnfp [Relativo, non determinante, feminino, plural]:

Senón se quere intervir directamente tamén se pode investir en empresas vinícolas que cotizan en bolsa. [CORGAetq: *ANT2002-10-17/29*]

Tampouco nesta cita se avanzou no financiamento autonómico nin nas reformas estatutarias, dúas cuestións espiñentas nas que o PSOE preferiu manter a liña da negociación multilateral exposta no programa das últimas eleccións. [CORGAetq: *ANT2008-07-10/74*]

Esta medida de apoio ás familias galegas únese a outras que xa figuraban outros anos na mesma lei, como a deducción fiscal de 20.000 pesetas polo nacemento dun fillo durante o ano 2001. [CORGAetq: *CG2000-11-22/3*]

Neste caso a sala é pequeneira pero ampla, moi apta para actuacións musicais de esas que merecen ser escoitadas en silencio. [CORGAetq: *GH2007-03-30/112*]

Tnms [Relativo, non determinante, masculino, singular]:

O produtor oferta o viño en xaneiro, dous anos antes de saír ao mercado, a un prezo que hai que pagar por adiantado. [CORGAetq: *ANT2002-10-17/29*]

O programa A Revista conectará esta xornada en directo co Parlamento de Galicia, xa que se conmemora o 75 aniversario do voto feminino, unha celebración que terá como acto central a representación teatral do proceso polo que en 1931 se aprobou o sufraxio feminino. [CORGAetq: *GH2007-03-30/99*]

Entrementres o home anda por aí alcumando de xeito choqueiro a todo aquel que non comulga coas súas rodas de muíño envelenadas. [CORGAetq: *ANT2005-03-03/6*]

Tnmp [Relativo, non determinante, masculino, plural]:

De feito, dos cinco concelleiros que somos, tres temos esa tendencia. [CORGAetq: *ANT2005-03-03/19*]

A axencia terá unha dotación inicial de mil millóns de pesetas pero está previsto que os fondos cos que conte non sexan inferiores ós 30.000 millóns nos próximos seis anos. [CORGAetq: *CG2000-11-22/3*]

Pois esa garantía debe ser para a totalidade dos traballadores, e non para uns cantos que lle puideran ser afíns a CC OO. [CORGAetq: *ANT2005-03-03/54*]

Se no antecedente non son pertinentes os atributos de xénero e número, como ocorre por exemplo cos patronímicos e topónimos ou cos adverbios, fixémonos na

función sintáctica que desempeña o relativo na cláusula incrustada, e se esta é a de Suxeito, por mor da concordancia que debe establecerse en xeral entre este e o seu Predicado, asignamos o valor específico *singular* (s) ou *plural* (p), segundo corresponda, e mantemos a ambigüidade no resto de casos. Así se explican as seguintes etiquetaxes:

Tnas [Relativo, non determinante, masculino/feminino, singular]:

Lagares Calvo, **que** ofreceu onte unha conferencia sobre 'A corporación entre as caixas españolas', dentro das xornadas sobre 'As caixas ante o terceiro milenio', salientou que cada día estas entidades teñen máis clientes, máis depósitos e ofrecen máis créditos cós seus competidores. [CORGAetq: CG1999-05-20/57]

Tras esta rebaixa, Cepsa, **que** controla o 25 por cento do mercado de distribución de hidrocarburos, comercializa a gasolina eurosúper a un máximo de 136,9 pesetas, a súper a 146,9 pesetas, a sen chumbo de 98 octanos a 148,9 pesetas e o gasóleo de automoción a 125,9 pesetas litro. [CORGAetq: CG2000-12-21/43]

Ademais, Carta Blanca inclúe o tema en directo Foliada das Crechas, **que** rende homenaxe á Casa das Crechas. [CORGAetq: ANT2008-07-10/48]

Tnap [Relativo, non determinante, masculino/feminino, plural]:

Nin tampouco relacionar a súa proposta coas de Euskadi e Cataluña, **que** por moito que sexan diferentes forman parte da mesma e inseparábel conxuntura política. [CORGAetq: ANT2005-03-03/5]

Fíxeme lurófilo por influencia dos chocos en tinta e das luriñas enceboladas **que** comía na casa e, inevitablemente, tamén por razóns literarias: [CORGAetq: GH2007-03-30/89]

Hai máis artistas **que** recrean isto que logo fotografan, mais o interesante no seu caso é a reviravolta sociopolítica que lle imprime aos motivos fotografados. [CORGAetq: ANT2008-07-10/45]

Tnaa [Relativo, non determinante, masculino/feminino, singular/plural]:

"É evidente que onde queira **que** se cree un baleiro, a natureza fai que se encha doutra maneira, e xa en Galicia temos algúns casos tamén doutras emigracións", dixo. [CORGAetq: CG2000-12-29/10]

Agora **que** comeza a tempada de presentación de resultados do primeiro semestre de moitas empresas, poderemos avaliar, con coñecemento de causa, o impacto real da crise nas contas das compañías. [CORGAetq: ANT2008-07-10/72]

Cultura, historia, relax e o verde das paisaxes **que** nos ofrecen as fermosas Terras de Celanova
[CORGAetq: *GH2007-03-30/106*]

Como conxunción subordinante non ten valor anafórico e non desempeña ningunha función no interior da cláusula incrustada; só serve de nexos para, entre outros, introducir unha subordinada substantiva en función de Suxeito, Complemento directo, Complemento Preposicional ou termo dunha Frase Preposicional que modifica a un substantivo, adxectivo ou adverbio. Vexamos algún exemplo de cada:

Introdutor dunha subordinada substantiva en función de Suxeito:

Con todo, é verdade que resulta difícil avaliar o sesgo introducido por estes factores, máxime cando poden, en moitas ocasións, poñerse exemplos de maiores custos ligados á acción de factores antitéticos aos anteriores. [CORGAetq: *GI1996-05/55*]

É ilusorio que o deflactor do PIB se sitúe en 2,8% cando estamos nunha inflación subxacente de 3,6% real. [CORGAetq: *ANT2002-10-03/32*]

Sería desexábel que a lei fose acompañada polo compromiso entre os grupos parlamentarios de manter un incremento sostido dos fondos adicados á cooperación internacional, e que este acordo se extendese ás outras administracións públicas galegas, nomeadamente as locais. [CORGAetq: *ANT2002-10-31/5*]

Introdutor dunha subordinada substantiva en función de Complemento Directo:

Dende esta institución aseguran que a demanda de madeira triplicarase nos próximos 30 anos. [CORGAetq: *ANT2002-10-10/33*]

O Financial Times afirma que o Brasil vai camiño de ser "superpotencia". [CORGAetq: *ANT2008-07-10/81*]

Así, subliñou que o ferrocarril tamén lle pasa "ao pé da porta" a Ribadeo e, por iso, lamentou que as autoridades da Xunta "non estivesen tan oportunos para ter isto previsto" e ofrecer unha alternativa distinta á construción dunha terminal flotante de carga e descarga que a empresa Galigrain proxecta construír no porto de Ribadeo. [CORGAetq: *GH2008-08-05/8*]

Introdutor dunha subordinada substantiva en función de Complemento Preposicional:

Aparicio insistiu onte nunhas xornadas en que os que contribúen son os cidadáns [CORGAetq: *CG1999-05-14/45*]

A OCU informa de que un 12% das gasoleneiras ten unha desviación de, alomenos, un 1% [CORGAetq: *CG1999-05-26/50*]

"Soño con que o Liceo volva crear unha hexemonía no hóckey, e creo que imos ben encarrilados" [CORGAetq: *DLV2007-05-15/3*]

Termo de frase preposicional modificando a un substantivo:

Deixemos á beira o argumento, mellor dito, a palinodia de que os cataláns son uns fenicios e uns aproveitados e, deixemos tamén á beira, a andrómena de que o novo sistema non nos afecta porque Galicia terá compensacións, garantías e vantaxes varias. [CORGAetq: *GI1996-06/41*]

Se di que si crerei o presidente do Goberno e se non di nada crerei que oculta algo, pero os navarros teñen dereito a que esa pregunta sexa respondida", subliñou. [CORGAetq: *GH2007-03-17/63*]

En apoio de que se protexa este patrimonio, téñense organizado varias marchas veciñais de protesta, demandando a consideración destes espazos como ben de interese cultural e a súa inclusión na Rede Natura. [CORGAetq: *ANT2008-07-10/59*]

Termo de frase preposicional modificando a un adxectivo:

"Os traballadores deberían estar moi conformes con que a compañía teña posibilidade de crecemento e de conserva-las súas fontes de traballo", sinalou. [CORGAetq: *CG2001-10-04/35*]

Non obstante, non está disposto a perder puntos do seu carné por iso, polo que recorrerá a sanción e xa consultou un asesor que se mostrou confiado en que a multa non chegará a máis. [CORGAetq: *GH2007-03-17/40*]

Son conscientes de que coñecer idiomas é esencial [CORGAetq: *GH2008-08-05/30*]

Termo de frase preposicional modificando a un adverbio:

O pasado domingo, o presidente do club, Giorgio Pedraneschi, rompeu o trato logo de que os ingleses amosasen a súa inquietude sobre unha suposta lesión de xeonllo do colombiano e pedisen unha redución do traspaso, fixado en ó redor de 1.250 millóns de pesetas. [CORGAetq: *CG1996-02-08/116*]

Faríaos catorce anos despois de que se lle impuxese a prohibición de construír buques. [CORGAetq: *TNI997-04/7*]

Antes de que se produza terá que pasar polo filtro da competencia. [CORGAetq: *CG2000-11-11/68*]

A conxunción subordinante *que* introduce, así mesmo, o segundo termo dunha estrutura comparativa e o dunha estrutura consecutiva. Por exemplo:

a) Segundo termo da comparación

Ao respecto, lembrou que os dous anteriores plans, de 1983 e 1995 _cando gobernaba o PP_ "igual que agora foron suspendidos a dous meses dos comicios". [CORGAetq: GH2007-03-17/10]

Os datos revelan que a ralentización da economía galega é menor que a española quizais debido a que estaba empezando un despegue do sector do ladrillo que, aínda así, estivo a piques de destruír para sempre o litoral. [CORGAetq: ANT2008-07-10/71]

Aquí sobra talento, pero a animación ten un proceso máis longo e máis custoso que a ficción, e necesita delas para visualizarse. [CORGAetq: ANT2008-07-10/23]

b) Consecuente

Así que temos unha situación tal que neste outono entraremos nunha fase de incertidume e non se pode calcular o que vai pasar máis alá. [CORGAetq: ANT2002-09-12/34]

O sentimento de pertenza é tan importante que reina por riba da razón, aínda que se trate de grupos onde a razón debería de ser o máis importante. [CORGAetq: ANT2008-07-10/9]

A axenda estival da tele ten tanta oferta, tanta competencia... que o anormal é ser atención das cámaras un ano si e outro tamén. [CORGAetq: GH2008-08-05/107]

En menor medida, aparece tamén como introdutor da Causa dunha estrutura causal:

"De reconversións nada _di Xosé Díaz_, que a min non me consta que se abrisen empresas nas que recolocar os traballadores dos estaleiros". [CORGAetq: TN1997-04/7]

A Fábrica de Moeda e Timbre saca a última tirada de billetes de dez mil, malia que aínda se acuñarán moedas, que solto sempre fai falla. [CORGAetq: CG2000-11-22/80]

Fomentar o diálogo a nivel nacional e internacional e confiar nos novos artistas con proxectos interesantes, dándolle pé para realizar novos traballos, que non todo se acaba no cubismo. [CORGAetq: ANT2005-03-03/71]

É, ademais, o nexo que une os auxiliares *haber* e *ter* cos infinitivos auxiliados conformando dúas perífrases modais que expresan obrigatoriedade:

a) Haber que + infinitivo

Habería que darlle maior alegría á economía", sinalou. [CORGAetq: CG1996-02-16/101]

De non ser así, haberá que avanzar no proceso de privatización de servizos públicos retirando de facto o exercicio de competencias ás institucións locais. [CORGAetq: ANT2002-10-17/30]

"Hai que ver se é posible a reconciliación entre o groso do pobo e unha mancha de ladróns lerchas que o saquearon". [CORGAetq: GH2008-08-05/45]

b) Ter que + infinitivo

"Temos que asumilo _sinalou_ como tamén temos que asumi-la axuda ó Terceiro Mundo, e que as empresas galegas teñen que busca-lo financiamento polos seus propios medios". [CORGAetq: CG2000-12-29/17]

Así, o vicepresidente da Xunta, Anxo Quintana, tivo que esforzarse onte para chimpas polas pedras que conducen á aldea de Saa, no municipio ourensán de Vilar de Santos. [CORGAetq: DLV2007-05-15/22]

Alguén ten que contar a verdade! [CORGAetq: ANT2008-07-10/27]

Finalmente, é usual, así mesmo, o emprego da conxunción subordinante *que* como reforzo da afirmación con *si*:

_Profesionalmente, si que son bastante ambicioso. [CORGAetq: GI1996-06/30]

É imposible chegar ás 30 toneladas de uva por hectárea con Treixadura, pero si que se pode alcanza-la metade, o que a 400 pesetas sería de 6 millóns por hectárea. [CORGAetq: CG2001-10-04/5]

Se alomenos gaña tanto como perde, se engade algún aspecto novo, entón si que a fago. [CORGAetq: ANT2005-03-03/49]

Pode resultar estraño ver no sistema de XIADA un *que* etiquetado como conxunción coordinante cando, en xeral, na tradición galega non se recolle (RAG/ILG, 1995 e 2003; Álvarez, Regueira e Monteagudo, 1986; Hermida, 2004) ou non se delimita como tal (Álvarez e Xove, 2002¹⁸⁶); salvo Freixeiro (2000: 582), quen si especifica o valor coordinante copulativo, mais faino só para as “*construcións intensificadoras que repeten a mesma forma verbal*: Dar, dá que ver a mar aborrecida,/ola vai e ola vén, brúa **que** brúa”.

Tan infrecuente é este uso, que non saltou ningunha ocorrencia no corpus de adestramento da prensa e tivemos que agardar ao relato para documentalalo:

Sen embargo non intimaron ata que unha vez, na trastenda, o profesor, ulisca que ulisca, dun libro a outro, foi bater co fociño nas partes do libreiro, onde fozou con aspiracións nasais gorentosas, ante a pasividade, a sorpresa e un certo deleite contido deste último. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

¹⁸⁶ Que forma parte de perífrases verbais (ter que + Part., haber que + Inf.) e de construcións de carácter adverbial formadas pola repetición dun verbo na súa forma de Imp. (corre que corre, xunto con corre corre) ou con Ind. Fut. no segundo membro (corre que correrás). [Álvarez e Xove, 2002: 650-651]

Etiquetamos tamén como conxunción coordinante o *que* presente nas construcións nas que, coa súa axuda, se combinan os indefinidos *algún* e *outro*, denominadas locucións determinativas ou pronominais na tradición española, mais practicamente ignoradas na galega:

Freixeiro (2000: 255-256)

A coordinación copulativa de *un e outro* en plural adquire un valor de totalidade, equivalente a 'todos' (*Uns e outros son iguais*):

*Non hai dúbida de que **uns** e **outros** impiden o rexurdimento de Hespaña* (Castelao, SEG 159)

Unidos pola partícula *que* conseguen un significado equivalente ao cuantificador indefinido *algún* (*Temos un que outro piso alugado*):

*O carrolo saille para fóra; na calva ten **unha** que **outra** serda* (Risco, PDP 11)

Álvarez e Xove (2002: 489)

Coma os máis cuantificadores, *algún* pode ir seguido de *outro*, achegando cada determinante o valor que lle é propio. Pero *algún* pode combinar tamén con este identificador nunha construción con *que*, que ten toda ela o valor de *algún*, só que destacando a escaseza da cantidade; en lugar de *algún* tamén se usa *un*:

¿Liches *algunha* *outra* novela dela?

Achegouse por alí *algún que outro* a botarlles unha man. Aínda se volveu dar *algún que outro caso* ~ *algún caso que outro* [dobre posibilidade de orde]

Díxome que fumara *unha vez que outra* ~ *algunha vez que outra* ~ *algunha que outra vez*.

Dous son os factores que nos condicionaron para tomar esta decisión. Por unha banda, o feito de non ter delimitado no etiquetario locucións en categorías variables; pola outra, malia non documentala aínda no corpus de adestramento, a imposibilidade de tratar computacionalmente a discontinuidade que amosa esta estrutura nunha das súas variantes:

pero agora xa non son unha forza, por máis que poidan facer aínda algunha que outra barbaridade. [CORGAetq: ANT2008-07-10/84]

Dende aquela teño pasado na compañía deles algúns dos mellores momentos da miña vida, e corrido algunha que outra farra. [CORGAetq: E. R Ruibal, *De corpo enteiro*, Xerais, 1994]

Ningún sabe de que se trata e entón eu vou e explicollo coa maior gracia de que son capaz, sempre servíndome, á par das miñas palabras, de movementos da man pequena con algún que

outro aceno chusco que para o caso teño ben ensaiado. [CORGAetq: E. R Ruibal, *De corpo enteiro*, Xerais, 1994]

Quizais algunha que outra vez, causará tamén lástima, quen sabe... [CORGAetq: E. R Ruibal, *De corpo enteiro*, Xerais, 1994]

A esa altura da súa vida, xa se fixera tamén cunha colección de arte, do "máis estimábel", formada con pezas de artistas galegos e españois, anque tamén de algúns que outros americanos e europeos. ... [CORGAetq: E. R Ruibal, *De corpo enteiro*, Xerais, 1994]

Algún que outro ocioso turista arredábase ás veces da autoestrada coa teima de coñece-lo vello casco cidadán e entretíñase logo nalgunha cafetería, uns minutos tan só, pra marchar axiña e acaso xa non regresar nunca. [CORGAetq: A. Conde, *A casa de Adara*, Xerais, 1996]

fronte a, por exemplo

Pode que lle beba algunha "cruzcampo" que outra, máis nada... [CORGA: A. Álvarez Cáccamo, *Xente de mala morte*, Galaxia, 1993]

E non son tan parvo como di a xente, que porque non traballo nin fixen familia e porque me gusta mollar algo a palleta, algún ribeiriño que outro se hai quen invite, pois xa pensan que un é parvo. [CORGA: S.uso de Toro, *Tic tac*, A Nosa Terra, 1993]

Empezou a deixarse caer por algunha verbena que outra. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

A rapaza atopábase de vacacións na casa dunhas curmás e algunha vez que outra viñeran os dous en moto a pasar o día neste areal. [CORGA: X. M. Martínez Oca, *Como paxaro caído do niño*, Baía, 2006]

Polo que respecta ao *que* interrogativo-exclamativo, pouco é o que pode diferir o seu tratamento en XIADA do de calquera gramática, salvo a aplicación nel dos atributos *xénero*, *número* e *valor*.

Como xa comentamos anteriormente en varias ocasións, malia que por si só non presenta variación xenérica, numérica nin funcional, o feito de traballar en lingüística de corpus fai que apreciemos sobremaneira os elementos que conforman un contexto dado e, sobre todo, aqueles que constitúen unidades diferenciadas, aínda que sexan menores, coma as frasais. En xeral, a aplicación do *valor* é a única licenza que concedemos polo momento á sintaxe no sistema XIADA. O funcionamento dos atributos antes mencionados non difire en absoluto da súa utilización, por exemplo, nos demostrativos, só hai que ter en conta as diferenzas obvias entre unha clase e a outra. Así, o interrogativo-exclamativo pode acompañar un adverbio, co que é de aplicación o valor

determinante a pesar de que aquí lle acaia mellor o significante *modificador*; pola contra, cando non forma unidade con ningún outro elemento caracterizámolo como *non determinante*, ou sexa *núcleo*. Algo semellante sucede cos valores de *xénero* e *número*:

1) Cando acompañan un *token* en cuxa definición son pertinentes ambos os valores, o interrogativo-exclamativo adopta eses mesmos valores:

Gdfp [Interrogativo-exclamativo, determinante, feminino, plural]

Que particularidades ten? [CORGAetq: ANT2005-03-03/49]

Que distintas as imaxes que nos dan a serie e a película, quizais porque, como di Tamar Novás, tentouse tratar con fidelidade e con respecto a nosa realidade, fuxindo de prexuízos extemporáneos e o resultado foi este, a admiración por unha historia real que nos dá a exacta dimensión da vida. [CORGAetq: ANT2005-03-03/35]

Gdfs [Interrogativo-exclamativo, determinante, feminino, singular]

¿Que opinión lle merece a reforma do mercado laboral? [CORGAetq: CG1999-05-09/9]

O que me pregunto despois é que condenada razón impulsou ao presidente dos kazakos a vir aquí, co de problemas que cheira ao redor! [CORGAetq: GH2008-08-05/54]

Gdmp [Interrogativo-exclamativo, determinante, masculino, plural]

Que criterios segue para escoller cada programa? [CORGAetq: ANT2005-03-03/49]

"Descoñezo en que termos Pablo Casado se referiu ao apartheid lingüístico, mais temos contacto permanente; [CORGAetq: ANT2008-07-10/12]

Gdms [Interrogativo-exclamativo, determinante, masculino, singular]

O non irlandés é unha invitación a encetar un debate autenticamente político sobre que tipo de Europa desexamos". [CORGAetq: ANT2008-07-10/19]

Que papel deben xogar as televisións? [CORGAetq: ANT2008-07-10/23]

2) Sempre que o interrogativo é *non determinante* e desempeña a función de Suxeito aplicámoslle o número da forma verbal coa que conforma cláusula:

Gnas [Interrogativo-exclamativo, non determinante, masculino/feminino, singular/plural]

Resulta difícil coñecer que tería ocorrido sen eles pero, en todo caso, esa constatación ilústranos sobre a súa insuficiencia para contrarrestar a acción dos mecanismos de mercado nun contexto de crise. [CORGAetq: GI1996-05/55]

Que quedou daquilo? [CORGAetq: ANT2005-03-03/62]

3) Se aparece no interior dunha predicación atributiva e o seu correlato amosa xénero (Suxeito-Atributo), caracterizámolo co mesmo valor que este presenta:

Gnms [Interrogativo-exclamativo, non determinante, masculino, singular]

Discutir que é un clásico. [CORGAetq: ANT2005-03-03/18]

Se fose elixido alcalde da Coruña, ¿que sería o primeiro que faría? [CORGAetq: DLV2007-05-15/20]

4) Cando acompañan unha unidade para a que non son pertinentes os atributos de xénero e número, ou son nucleares mais non temos elemento ningún no contexto que nos permita concretar eses valores, aplícase a non definición; é dicir, mantense a ambigüidade na posición do xénero e na do número (*masculino/feminino, singular/plural*):

Gdaa [Interrogativo-exclamativo, determinante, masculino/feminino, singular/plural]

"Qué máis quixesemos nosoutros", sinalou o presidente de Erimsa. [CORGAetq: CG1996-02-07/38]

Gnaa [Interrogativo-exclamativo, non determinante, masculino/feminino, singular/plural]

¿Que votaría vostede se mañá houberse un referendun sobre a permanencia do seu país na UE? [CORGAetq: TNI997-04/7]

Que significa este premio para a súa carreira? [CORGAetq: ANT2005-03-03/49]

Naturalmente, sabemos que estamos concedendo un significado duplo aos valores *aa* na posición de xénero e número (terceiro e cuarto lugar da cadea de etiquetaxe, respectivamente). Por unha parte, utilízanse cando nos resulta imposible determinar o xénero e número do substantivo ou adxectivo que acompañan e, pola outra, empréganse para caracterizar o interrogativo-exclamativo cando este conforma unha unidade independente ou con outro elemento no que non aplica nin o xénero nin o número. Mais é unha peculiaridade do *que* interrogativo-exclamativo no que cremos que a exhaustividade de análise compensa a duplicidade significativa.

Contabamos con que se manifestaran no subcorpus de adestramento ficticio as etiquetas Gdap, Gdas, Gnap, Gnfs, Gnfp e Gnmp; non obstante, cando levamos xa desambiguado e revisado preto da metade dese conxunto, aínda non emerxeu ningunha delas. Tanto é así, que temos que seguir botando man para a súa exemplificación do *kernel* ou núcleo de XIADA, ou ben recorrer a outros membros do paradigma dos Interrogativos-exclamativos¹⁸⁷:

Gdap [Interrogativo-exclamativo, determinante, masculino/feminino, plural]

¡Que fáciles! [kernel]

Gdas [Interrogativo-exclamativo, determinante, masculino/feminino, singular]

Que terrible. [kernel]

Gnap [Interrogativo-exclamativo, non determinante, masculino/feminino, plural]

¿E quen veñen á Mesa de Programación? [kernel]

Gnfp [Interrogativo-exclamativo, non determinante, feminino, plural]

A deputada presentou unha serie de preguntas ante a Cámara galega co fin de que o conselleiro de Sanidade, Hernández Cochón, explique cáles foron as causas e os criterios para lle adxudicar á clínica privada este convenio. [kernel]

Cales son as alternativas? [CORGAetq: ANT2008-07-10/70]

Gnfs [Interrogativo-exclamativo, non determinante, feminino, singular]

Cál é a vosa representación como delegación da seguridade do pobo, ilustres deputados do PSOE. [kernel]

¿Cal é a día de hoxe a situación do galego? [CORGAetq: DLV2007-05-15/62]

Gnmp [Interrogativo-exclamativo, non determinante, masculino, plural]

"Vimo-las mesmas imaxes que vostedes na televisión, pero non estamos sobre o terreo e non coñecemos-la identidade dos pasaxeiros, nin sabemos cántos poden ser", engadiu. [kernel]

Cantos dos espectadores que vén estes días a película ignoraban que o Berlín nazi foi tomado polos soviéticos e non polos norteamericanos? [CORGAetq: ANT2005-03-03/51]

¹⁸⁷ Malia que as catro últimas poidan resultar esporádicas, previamos que na ficción se desen exemplos variados das dúas primeiras, pois o estilo de lingua (enunciados curtos, diálogo etc.) así o facía supoñer. Porén, ou ben esas etiquetas son moito menos frecuentes do que criamos ou ben o azar ao adoptar a forma de extracción aleatoria de parágrafos, método escollido para a selección do subcorpus ficticio, non estaba polo labor de botarnos unha man.

Do mesmo xeito que non aceptamos que se nos diga desde Madrid con quen pactamos, tampouco desde Cataluña dicimos cales deben ser os pactos", argumentou. [CORGAetq: GH2008-08-05/61]

4.4.5. Categoría Verbo

4.4.5.1. Participio vs. adxectivo

A etiquetaxe de grande parte das formas ortográficas remisibles tanto á categoría Adxectivo como á categoría Verbo, na que se integra a forma verbal non persoal Participio, é problemática. Partindo desta premisa e tendo en conta que o obxectivo que perseguimos é proporcionar unha etiquetaxe comprensible e predicible para a maior parte de usuarios do noso sistema, comprenderase que non tratemos de afondar nos criterios de delimitación destas formas e optemos, simplemente, por seguir unha liña xa marcada. O marco teórico do que partimos para establecer a diferenciación é o recollido en Álvarez, Regueira e Monteagudo (1986), resumido xa no parágrafo introdutorio da epígrafe pertencente ao Participio:

Polas súas características morfolóxicas e o seu comportamento sintáctico, o participio atópase máis preto do adxectivo ca do verbo. De feito, soamente ten valor plenamente verbal nas perífrases verbais (v. p. 404) e nun limitado número de construcións. [Álvarez, Regueira e Monteagudo, 1986: 398-402]

Así, os supostos nos que remitimos á forma verbal participio son os seguintes.

1) Construcións absolutas:

Dito isto quixera formular tres cuestións: [CORGAetq: ANT2005-03-03/24]

Dada a obamanía do zapaterismo podemos saír da crise económica e facernos ricos revendéndolles os produtos da Obamastore aos seguidores de Z. [CORGAetq: GH2008-08-05/96]

O mobiliario, o enxoval e o resto do contido de "Hyperhábitat" dispón dun microordenador que permite o seu "movemento" grazas a un "slider" (dispositivo de control) e que recollerá todas as propostas dos participantes nun rexistro que, unha vez acabado o proxecto, se analizará. [CORGAetq: GH2008-08-05/50]

A penas aprobada a lei, creáronse bancos de emisión en Bilbao, Santander, Málaga, Sevilla, A Coruña e outras cidades, sempre por decisión de comerciantes e empresarios locais. [CORGAetq: CG1996-02-09/57]

2) Construcións pasivas:

Este grupo foi fundado por estes dous mecenas cunha finalidade: [CORGAetq: GH2008-08-05/94]

A primeira parte da obra publicouse no 1973, en París, ata onde chegou grazas á súa secretaria, que acabou suicidándose despois de ser capturada e torturada polo KGB. [CORGAetq: GH2008-08-05/44]

Por iso, o parlamentario do PPdeG incidiu en que se converte o "Estado de Dereito na vía dos feitos", o que foi negado polo director xeral. [CORGAetq: GH2007-03-17/19]

A medida que pasa o tempo aumenta a desesperación dos pais de Madeleine McCann, a nena británica de catro anos que foi secuestrada no Algarve luso mentres os seus pais estaban nun restaurante a apenas 50 metros da casa. [CORGAetq: DLV2007-05-15/69]

3) Perífrases verbais

3.1.) DAR + PARTICIPIO.

O Celta non dá levantado a cabeza e ten cada vez máis difícil a salvación. [CORGAetq: DLV2007-05-15/103]

"Choveu continuamente e o río non deu tragado con todo, quedamos sen mobles e agora ninguén paga as indemnizacións", indican esta parella e Finuca, outra afectada. [CORGAetq: GH2007-03-17/6]

De ser así, non dou collido o sono e, máis tarde ou máis cedo, acabo por decatarme. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

3.2. LEVAR + PARTICIPIO.

Canal Plus é instrutivo pero incluso Redes _do relambido Punset_ xustificaría ao teleadicto, pois difunde noticias sensacionais, como por exemplo que cada dez anos cambian tódalas células do noso corpo, de maneira que un tipo de 48 tacos, coma min, xa leva habitado catro carrocerías e está a piques de estrear unha quinta. [CORGAetq: ANT2005-03-03/9]

Polo que levamos visto, os organizadores do festival superaron con tranquilidade as primeiras probas. [CORGAetq: ANT2008-07-10/24]

O zapateiro estivo un bo cacho facéndose o interesante, pagou o que xa levaba gastado no seu e nas rondas ofrecidas ós presentes, e ó pagar botou a man a un peto calquera e sacou un feixe de billetes de mil. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

3.3. TER + PARTICIPIO.

Tamén sinala que "aínda non temos pensado onde instalar as plantas de produción nin quen será o fornecedor de material xa que hai que ir pouco a pouco". [CORGAetq: ANT2002-09-12/29]

Pero ademais ten abordado nos últimos anos diversos proxectos audiovisuais e expositivos, continuando unha senda que xa abrira nos anos 60 en colaboración coa discográfica Edigsa, que puxo en circulación os discos inaugurais do movemento Voces Ceibes. [CORGAetq: ANT2008-07-10/4]

Ten afirmado Vargas Llosa _tan contradictorio en tantas ocasións_ que a literatura non é unha actividade gratuíta: [CORGAetq: GH2007-03-30/86]

3.4. HABER + PARTICIPIO

O galego rexeita o emprego das formas do verbo *haber* con participio formando unha perífrase. Cando este verbo aparece cun participio, normalmente está en forma impersoal e o participio concorda co complemento directo como complemento predicativo del. Por exemplo:

Ó primeiro vistazo, xa dei un repaso ó que alí había espaxado: un váter medio partido, un prato de porcelana e unha bicicleta escacharrada; todo moi cotroso, lóbrego. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Deste xeito, antes de Semana Santa habera instalados 300 radares fixos nas vías, e outros 220 móbiles. [CORGA: GH2007-03-13/26]

En Galicia hai catalogados medio cento de vehículos históricos [CORGA: CG2000-12-10/15]

Non obstante, pode aparecer ocasionalmente a perífrase formada con *houbera* ~ *houbera* + *participio* co valor temporal de antepretérito e modal de irrealidade:

Ogallá se houbera namorado entón, alí mesmo, mentras él sacudía o cabelo húmido e sorría picareiro. [CORGAetq: C. Pereiras, *pequenos infinitos*, Difusora de Letras Artes e Ideas S.L., 2008]

A hipótese consiste en proclamar que, se Valle houbera redactado parte das novelas e dramas en idioma galego, a situación da lingua sería agora diferente. [CORGA: GH2005-02-27/33]

O curioso é que estas formas de construción, de remota estirpe árabe, estas paredes azulexadas proveñen de Portugal e houberan podido reproducilas con só cruzar o Miño. [CORGA: CG2000-12-03/17]

Sacados os tres apartados anteriores, reflexo dos contextos situacionais prototípicos unívocos das formas verbais participiais nos que non pode concorrer un adxectivo, no resto dos casos, dada a proximidade entre uns e outros, o participio, igual ca o adxectivo, aparece principalmente modificando un substantivo ou funcionando

como atributo ou predicativo, ben do suxeito, ben do complemento directo. Os seguintes exemplos dan conta destes outros comportamentos funcionais:

4) Modificador de substantivo:

Deste xeito, o Plan Vigo Íntegra esquece as medidas propostas pola Unión Europea en canto á mellora da calidade do ambiente que obrigan aos estados membros favorecer a eficiencia enerxética, tamén na dotación das infraestruturas. [CORGAetq: GH2008-08-05/9]

Todos cren que é un agasallo enviado polo propio Obama, unha mostra do seu aprecio cara á forma actual do socialismo español, o zapaterismo. [CORGAetq: GH2008-08-05/96]

Mini e Mero saben moi ben que restituír significa normalizar, e normalizar equivale a continuar o destino trazado polos nosos devanceiros. [CORGAetq: GH2008-08-05/94]

Asegurou que o lema elixido polo seu partido para esta campaña, Centrados en ti, garantía de goberno, recolle a idea do PP de ofrecer un programa próximo ós cidadáns e ironizou sobre o lema Novos tempos do PSOE. [CORGAetq: GH2007-03-17/24]

5) Predicativo, ben do suxeito ben do complemento directo:

Hoxe presenta 'Látego de Algas' na Coruña, acompañado do editor, Miguel Anxo Fernán-Vello [CORGAetq: GH2008-08-05/46]

Despois tivo lugar a mesa redonda no Ateneo da capital ourensá, presentada polo humorista Xose Lois González O Carrabouxo, quen asegurou que se trata de homenaxe "necesario e bo", porque "aínda a calor da presenza de Peña na rúa do Paseo non se foi". [CORGAetq: GH2007-03-17/26]

O único vehículo implicado foi o Renault 19 Chamade, conducido por Elías A.A, de 80 anos e veciño de Montederramo, que sofre feridas de gravidade ao saírse da vía. [CORGAetq: DLV2007-05-15/6]

Os filipinos votaron onte nuns comicios lexislativos e locais enturbados pola violencia e a fraude e considerados unha proba para a presidenta, Gloria Macapagal Arroyo, que pretende consolidar a súa maioría. [CORGAetq: DLV2007-05-15/65]

6) Atributo

Por aquel tempo, o zapateiro coxo era xa coñecido polo vecindario como un pobre homiño trastornado, porque cada poucos días alguén o atopaba fozando nos buzóns de uso exclusivo para carteiros, borracho e desesperado, ou tentando entrar por un sumidoiro; [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

O primeiro sobre que chegara, ía xa para un ano, chamáralle a atención pola impersoal pulcritude con que estaba escrito o enderezo; [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

_Os teus lentes están ben gardados. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

Varias son as cuestións que se enlean do que levamos visto deste apartado. En primeiro lugar, o recoñecemento das perífrases verbais en XIADA; en segundo, a identificación de perífrases verbais con participio, e en terceiro, a etiquetaxe das formas coincidentes a adxectivos e participios que non están entre os tres primeiros supostos exemplificados máis arriba.

Polo que se refire ao primeiro dos puntos, o tratamento das perífrases verbais en XIADA, non existen como tal unidade complexa no noso sistema. Os complexos verbais son analizados secuencialmente de xeito independente, segundo a análise que lle corresponda a cada unha das formas; é dicir, non delimitamos perífrases verbais, senón que cada forma verbal leva unha análise independente de xeito que non figuran no lexicón nin se poden localizar no sistema de consultas mediante unha etiqueta concreta. Para atopar perífrases verbais no CORGAetq é preciso que o usuario trate cada un dos constituíntes como un elemento gramatical distinto, e realice as buscas cubrindo o campo do lema na primeira casilla e a etiqueta na segunda ou na terceira, dependendo de se existe elemento nexual ou non. Por exemplo, para localizar todos os casos da perífrase aspectual perfectiva formada polo auxiliar *vir* + *de* + *infinitivo* aconsellamos proceder como se amosa na seguinte imaxe:

The image shows a search interface with a dark red background. It contains several rows of input fields and checkboxes. The fields are arranged in a grid-like structure. The first row has four input fields labeled 'Forma1:', 'Forma2:', 'Forma3:', and 'Forma4:'. The second row has four input fields labeled 'Etiqueta1:', 'Etiqueta2:', 'Etiqueta3:', and 'Etiqueta4:'. The third row has four input fields labeled 'Lema1:', 'Lema2:', 'Lema3:', and 'Lema4:'. The fourth row has four input fields labeled 'Unidade1:', 'Unidade2:', 'Unidade3:', and 'Unidade4:'. Below these are two dropdown menus: 'Área temática:' with 'Todas' selected and 'Subárea:' with 'Todas' selected. Then there are three more dropdown menus: 'Desde:' with '1975' selected, 'Ata:' with '2009' selected, and 'Medio:' with 'Todos' selected. Below these are six checkboxes with labels: 'Corpo', 'Pé de foto', 'Nota', 'Titular', 'Resumo', and 'Acoutación'. All these checkboxes are checked. Below these are six more checkboxes with labels: 'Prólogo', 'Apéndice', 'Dedicatoria', 'Cita', and 'Encabezamento'. All these checkboxes are checked. At the bottom right are two buttons: 'Limpar' and 'Buscar'.

A segunda das cuestións é máis problemática dende o punto de vista teórico ca do práctico. Trátase da delimitación de perífrase verbal con participio, ou dito doutro xeito, en que casos o participio cando vai posposto a *levar, ter, haber, estar, dar* ou *ir*, por exemplo, conforma en realidade un complexo verbal e non resultan dous constituíntes diferentes nos que, malia pertencer ambos á categoría verbal, non funcionan conxuntamente, senón que o primeiro é o predicado e o segundo un predicativo, o cal só nun segundo nivel de análise actúa como predicado dunha cláusula de participio. Ao longo do apartado das perífrases verbais con participio descritas por Álvarez, Regueira e Monteagudo (1986) a información a este respecto é confusa; así, en *dar + participio* advírtese que “*pode darse a posibilidade de concordancia entre o adxectivo e o obxecto directo*” (1986: 412), correspondendo a esta explicación o exemplo seguinte:

O pobre do soldado non daba xuntados os cartos para lle pagar ó taberneiro.

A continuación, para *levar + participio*, volve incidirse en que “*o participio pode aparecer concordando co obxecto directo*” (1986: 412), para cuxo caso introducen este outro exemplo:

Xa levo lidas unhas trinta páxinas.

Finalmente, en *ter + participio* especificase que se o participio concorda co complemento directo, como predicativo del, non nos atopamos ante unha perífrase verbal, por exemplo:

Teño gardada unha caixa de puros para vostede.

Como podemos ver, esta é outra cuestión que merecería un estudo máis polo miúdo, non obstante, posto que na práctica non afecta á etiquetaxe do corpus de adestramento por sernos indiferente se forma parte dunha perífrase ou non, dado que de todas formas vai ser catalogado como participio, non imos pararnos máis nela¹⁸⁸. Limitámonos a deixar constancia da problemática e a indicar que no CORGAetq se etiquetan, coma nos exemplos que seguen, como participios:

A revista, dirixida polo teólogo e académico Andrés Torres Queiruga, naceu no 1977 e desde entón leva publicados 150 números co galego como lingua vehicular. [CORGAetq: GH2007-03-17/49]

¹⁸⁸ Un traballo bastante interesante neste sentido, malia que para castelán, é o de Di Tullio (2008: 99-125).

Xa levaban bimbadas entre todos varias botellas, cando un dos seus mellores amigos, o Nécora, albanel de profesión, levouno a un aparte, a xente cantando as de sempre, e preguntoulle de onde viña tan desprendido. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Por hoxe, vai servida. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

As actuacións van dirixidas á venda ambulante, e establecementos que ofrecen produtos falsificados, así como á localización das redes de comercialización. [CORGAetq: *CG2000-12-28/35*]

Naquela miña nenez fugaz e intemporal cadroume coñecer un pouco a don Álvaro o Romano, un vello viúvo de falares breves e saborosos, e mais a Pastorita, a súa neta, que tería estreados os dezaseis anos, cando eu andaría polos oito ou nove. [CORGAetq: X. M. G. Trigo, *Extintos básicos*, Xerais, 1998]

"Sufrimos un baixón moi importante na zona caribeña, que se pode centrar en torno ó 50 por cento, sobre todo debido a que a xente está esperando a ver qué é o que ocorre para decidirse a viaxar, e anula a reserva que tiña feita con anterioridade". [CORGAetq: *CG2001-10-18/16*]

Finalmente, entendemos que o termo non marcado na oposición entre adxectivo e participio vai ser o adxectivo, polo que, en caso de non encadrarse a forma nalgún dos tres supostos iniciais de arriba –construción absoluta, pasiva e perífrase– a etiqueta que lle corresponde por defecto é a de adxectivo. Naturalmente, a aplicación desta premisa afecta única e exclusivamente ao corpus cuxa etiquetaxe foi revisada manualmente, o CORGAetq, pois na etiquetaxe automática, é dicir, cando non hai supervisión manual, o etiquetador só escolle unha das etiquetas que lle constan no lexicón, e por máis que os verbos foron introducidos masivamente a partir de dous manuais (Díaz Regueiro, 1992 e Nicolás Rodríguez, 1993) non sucedeu o mesmo cos correspondentes adxectivos.

Agora ben, dito isto, é obrigado recoñecer que a etiquetaxe das formas que poden remitirse a calquera das dúas categorías, Adxectivo e Verbo, cando actúan como modificadores ou predicativos, é un dos puntos negros de XIADA dado que non conseguimos ser sistemáticos nin atopar criterios claros de delimitación.

4.4.5.2. Infinitivo vs. substantivo

Tradicionalmente acostúmase dicir que as tres formas non persoais do verbo relaciónanse, polo seu comportamento funcional, con outras tantas clases de palabras coas que manteñen certas semellanzas: o infinitivo co substantivo, o participio co adxectivo e o xerundio co adverbio.

Neste apartado referiremos unicamente as decisións tomadas á hora de diferenciar infinitivo e substantivo na práctica da etiquetaxe do corpus de adestramento. Partiremos da concisa mais, non obstante, moi práctica e precisa distinción que Álvarez, Regueira e Monteagudo reflicten na súa gramática (1986: 385-386). Estes autores distinguen tres graos de aproximación do infinitivo ao substantivo. O primeiro deles non representa ningún problema para nós, posto que os orixinais infinitivos xa están totalmente lexicalizados como substantivos, e mesmo constitúen xa entradas léxicas independentes no dicionario. Destes non faremos máis ca recoller algún exemplo de uso, dado que son xa auténticos substantivos. Unha mostra:

Ante a vosa mirada expuxen todo o sucintamente que era o meu deber, o cadro do estado e marcha desta sucursal, e pausados pero firmes progresos que realizou, obedecendo a aquel 'apresúrate lentamente' da sabedoría antiga. [CORGAetq: CG1996-02-07/65]

Era o cantar da miña nai que eu máis gozaba de ouvir: forza, graza, alegría e raiba para bailar. [CORGAetq: ANT2005-03-03/75]

Do mesmo parecer é a case totalidade dos dirixentes que encabezan as distintas organizacións que integran o BNG e dos membros da dirección. [CORGAetq: ANT2005-03-03/15]

O seguinte chanzo na aproximación do infinitivo ao substantivo prodúcese coa substantivación funcional. Aquí ten lugar a asimilación ocasional por parte do infinitivo aos valores dun substantivo nunha frase nominal. Neste caso o infinitivo acepta determinantes, adxectivos e frases preposicionais con *de* que o modifican, a semellanza do que sucede cos substantivos. Na anotación do corpus de adestramento estes serán para nós nestes contextos substantivos. Por exemplo, os seguintes:

Tíñame collido pola queixada e pousábame a cabeza na súa barriga, que subía e baixaba ao compás do seu alentar. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Tiña un estar lixeiro e suave, de maneira que moitas veces nin era consciente da diferenza entre sentila na casa ou estar só, non interfería nas miñas decisións cotiás. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Era ela, correndo cara a unha morte segura e cruel ou cara á disolución naquel bruar mortífero. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Acompañaba as súas palabras con ademáns do seu braciño, que o axudaban a expresarse, completando ou reforzando o sentido do seu dicir. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Xa nos dirán máis adiante como se detén o furioso avantar das mareas. [CORGAetq: C. Pereiras, *pequenos infinitos*, Difusora de Letras Artes e Ideas S.L., 2008]

Por fin vanse todos e Xelote volve quedar só, co manso rebulir da auga afondando nas canas molladas. [CORGAetq: C. Pereiras, *pequenos infinitos*, Difusora de Letras Artes e Ideas S.L., 2008]

Nas rúas baleiras, alumeadas a treitos polas luces alaranxadas dos farois, ecoaba o rítmico golpear dos seus tacóns gastados, coma toques de tambor que sinalaban a irrevogable sentenza de morte. [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

As palabras do fotógrafo e colaborador de Adegas Xan Muras explican o sentir xeral de todos os integrantes desta mostra itinerante que pretende "botar un cabo a esta loita, ofrecendo a nosa particular visión do problema". [CORGAetq: GH2007-03-17/12]

O último dos chanzos, o máis separado do substantivo, acádase cando se substantiviza un infinitivo –en realidade, o predicado dunha cláusula de infinitivo– ben só ou ben con algúns dos complementos que o poden acompañar, mais, á fin e ao cabo, complementos que son os característicos do esquema argumental do verbo en cuestión. Se nos fixamos nos exemplos que seguen e que son mostra da etiquetaxe deste último chanzo, o infinitivo como forma verbal que continúa a ser –V0f000, polo tanto– non perde, como si sucede no anterior subapartado, o poder construírse con complemento directo, circunstanciais, complemento de réxime, etc., chegando incluso a acompañarse de pronomes enclíticos. Vexamos algúns destes casos do CORGAetq:

Por moito que intento non lle podo explicar que o escribir representa para min o único experimento serio de me comunicar co próximo, de provocar nel a reflexión e a discusión. [CORGAetq: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

A concordia, o consenso, o buscar os puntos de unión, serán camiño obrigado para non saír da vía. [CORGAetq: GH2007-03-30/88]

Non sería prudente para min o cambiar a aposta". [CORGAetq: CG2002-05-15/17]

Así mesmo, o subdirector do Banco de España, Eduardo Galilea Ortega, afirmou onte que o proceso de reorganización e reestructuración das caixas de aforros, que propiciou o pasar de 75 entidades en 1990 ás 51 actuais, fixo que estas incrementen a súa produtividade e melloren o seu grao de eficiencia. [CORGAetq: CG1999-05-20/57]

Fontes da campaña de McCain sinalaron onte que o recorrer ás reservas estratéxicas de petróleo non é un substituto para un plan real para aumentar a oferta adicional de enerxía mediante novas perforacións e enerxía nuclear. [CORGAetq: GH2008-08-05/66]

Pero foi nese momento, ó ver as paredes exteriores da cadea, cando se lle puxo a dozura fervente no estómago e non era o noxo da resaca nin o ter fumado moito, senón a cara de lambalcuza do Perla, e o trancazo dos pasadores pola noite cando veñen pechar e logo o ferro do pasador, o pau do te que forma o pecho, cae martelando na porta metálica, esa secuencia de ruídos exacta e ben gravada na memoria, que lle facía subir a náusea deica a gorxa; [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

'Estamos pagando o traballar exclusivamente para o Estado' [CORGAetq: *ANT2005-03-03/54*]

Naquel volver ao país, Cabanillas coñece e asume o fío ideolóxico e de compromiso que está a agromar detrás do denominado agrarismo, un termo social, económico, de rebeldía e imbricación na realidade labrega que tería como expoñente fundamental a Basilio Álvarez (o cura de Beiro, parte esencial daquel sindicato denominado Acción Gallega) que estableceu un discurso político de concienciación e de dignidade que axiña cativou o pensamento e obra de creación doutros intelectuais, como Noriega Varela ou Cabanillas. [CORGAetq: *GH2007-03-17/78*]

O outro día vin o teu home na tele, moi interesante o que dixo, sentínome orgullosa del. Interesante, que significaba esta palabra gastada, e o seu sentirse orgullosa expresábase nas súas chamadas telefónicas a tódolos coñecidos, para extenderse sen previo aviso a falar do seu intelixente xenro. [CORGAetq: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Fronte ao infinitivo do segundo chanzo, que funcionalmente actúa como un substantivo, neste é posible, así mesmo, documentalo construído, por exemplo, cunha cláusula con *que* que actúa como complemento directo. Non apareceu polo momento ningún caso no corpus de adestramento, pero si os hai no CORGA:

Pero é perfectamente cristiano o percurar que as diferencias do mundo seian o máis pequenas posible. [CORGA: X. Bernárdez Vilar, *Un home de Vilameán. Anatomía dunha revolución*, Patronato da Cultura Galega, 1976]

Ben sabe vostede o amigos que eramos... quero decir, que somos, o seu tío e máis eu, e o pensar que poidera estar morto, atordíume de tal xeito que non me atrevín a entrar soio na súa casa. [CORGA: X. M. Martínez Oca, *Número de patente / Querido Brais / A figueira*, Ediciós do Castro, 1978]

á explicación deste último quizais contribuíse o pensar que no medio rural ao traballo asalariado sumábaselle con frecuencia outras actividades, agrarias ou mariñeiras. [CORGA: M. González Probados, *Movemento obreiro e socialismo*, Ediciós do Castro, 1983]

Tamén o esquecer que, 19 anos despois da aprobación da Lei de Normalización Lingüística, a situación é completamente distinta. [CORGA: *ANT2002-09-05/14*]

4.4.5.3. Infinitivo vs. infinitivo conxugado

O sincretismo formal existente entre as formas do infinitivo e as da primeira e terceira persoas do singular do infinitivo conxugado provoca que non sexa posible discriminar entre un e outro infinitivo.

Álvarez, Regueira e Monteagudo (1986: 395) destacan a influencia que factores estilísticos e expresivos teñen no emprego das formas conxugadas e da invariable, de xeito que en moitas ocasións se podería utilizar calquera das dúas e só a énfase ou a expresividade nos decide por unha delas.

De todos son coñecidas as normas de uso da forma invariable e mais as das formas variables, polo que non ten obxecto reiteralas aquí. Non obstante, esas regras non teñen sempre reflexo nos textos, e así é frecuente que non se empregue o infinitivo conxugado aínda que o suxeito deste sexa distinto do do verbo principal:

Este último factor é preocupante, xa que ao desaparecer os ingresos que fornecía o Imposto á Actividade Empresarial aos concellos, éstos botarán man do seu principal activo para obter ingresos: o chan. [CORGAetq: ANT2002-10-24/31]

Esta estratexia require recursos públicos e privados en educación e formación dos traballadores, investimentos en infraestruturas e bens de equipo e deseño de produtos e servizos de calidade, pero, tamén, require tempo e paciencia que, ás veces, resulta excesiva ao estar sometidos os gobernantes ao ciclo político-electoral e os dirixentes empresariais á presión bursátil e da conta de resultados. [CORGAetq: ANT2005-03-03/55]

Ou tampouco cando, aínda sendo o mesmo suxeito, a subordinada de infinitivo vai colocada antes do verbo principal:

"Mesmo ó chegar xa nos despelotamos todos ó dereito e alí eu me sentín xa sen remedio entre aquel par de pés de nocelo groso e unllas esborranheadas, goroludos e abondo garfos. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Ó dobrar unha esquina, chegamos a unha rúa estreita e empinada. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Nun momento dado cremos, inxenuamente, que poderíamos sustentar a distinción polo menos nun caso: cando o suxeito do infinitivo fose distinto do do verbo principal e estivese explícito. Mais tampouco nestas circunstancias se podería afirmar que estamos ante as formas conxugadas, pois a explicitación do suxeito non implica

necesariamente o emprego do infinitivo persoal, como se comproba ao contrastar os seguintes exemplos:

ó entrar el, saudábao cunha reverencia cortés e logo petaba no mostrador para que lle escanciasen viño sen demora. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Un merlo procura a mantenza entre a herba a carón da pedra que cobre a tumba de frei Xoán, pero escapa voando ó chegar eles. [CORGA: M. Aleixandre, *Tránsito dos gramáticos*, Xerais, 1993]

Así pois, o emprego usual no CORGA do infinitivo en circunstancias nas que se recomendan as formas conxugadas, xunto coa ausencia de desinencias específicas de número e persoa que nos permitan distinguir as formas persoais da primeira e terceira persoas do singular do infinitivo conxugado do non persoal, determinaron que no noso corpus non se rexistrese ningunha ocorrencia correspondente a elas e optemos sempre pola etiqueta caracterizadora da forma invariable: V0f000.

4.4.5.4. Verbos *poñer* e *pór* e derivados

Dado que o noso sistema esixe que toda forma sexa remitida a un único lema, atopámonos ante as formas comúns a *poñer* e *pór*, e por suposto tamén nos seus derivados, na tesitura de ter que escoller só un. A maiores, posto que unicamente se nos permite, por condicionamentos computacionais, ter en conta a secuencia na que aparece inmersa a forma en cuestión, consideramos que o termo non marcado neste par é o verbo *poñer*.

Nun momento dado pensamos que poderíamos ter como unidade de referencia para determinar a escolla dun destes dous lemas o texto en si, a noticia no caso dos xornais e revistas ou o relato curto no caso das coleccións de relato. Mais o feito de prever que o subcorpus de adestramento de ficción se conformase partindo de unidades menores –parágrafos– e non de relatos completos, así como que o etiquetador/lematizador nada máis ten en conta os elementos presentes nunha oración, a opción de escoller entre *poñer* ou *pór* segundo fose o lema empregado maioritariamente en cada un dos textos converteuse en inviable. Ademais, aínda que sería o lóxico, non hai garantías de que un mesmo autor recorra no mesmo texto sempre ao mesmo lema. Unha mostra desta variación témola nunha noticia desambiguada do xornal *Galicia Hoxe* do día 5 de agosto do 2008:

O sindicato aproveitou así a operación dada a coñecer onte, de detención dun veleiro con 3.800 quilos de haxix fronte ás costas da Coruña para pór de manifesto as condicións nas que realizan o seu traballo os efectivos deste corpo. [CORGAetq: GH2008-08-05/11]

Ademais, a central sindical sinala que o escáner do porto de Vigo para o control e entrada de mercancías prohibidas en colectores non se puxo aínda en marcha, a pesar de que leva un ano instalado. [CORGAetq: GH2008-08-05/11]

Por iso, a CIG denunciou onte a existencia de "numerosas trabas" que se lle poñen aos funcionarios de Vixilancia Aduaneira para desenvolver o seu traballo na nosa comunidade, polo que se están reducindo as aprehensións en Galicia, mentres que a droga segue entrando como hai anos. [CORGAetq: GH2008-08-05/11]

Como vemos, o mesmo autor utiliza dúas formas remisibles moi claramente a cadanseu infinitivo: a da primeira secuencia a *pór* e a da última a *poñer*. Non obstante, tendo en conta os datos anteriores, a que lema debemos remitir *puxo*? Que criterio seguiríamos neste caso? A única opción sería un arbitrario.

Así pois, as formas comúns ao par *pór* e *poñer* e mais aos seus derivados remítense sempre no corpus de adestramento a *poñer* ou derivados.

4.4.6. **Claro, claro que...: interxección?**

Outra das cuestións que suscitou en nós algunha que outra dúbida foi a concorrencia de *claro* en contextos nos que non funciona nin como adxectivo nin como substantivo, usos estes últimos nos que non entraremos por considerar que non presentan ningunha particularidade especial.

Tampouco entraremos na categorización adverbial de *claro* en construcións do tipo *falar claro*, onde se aprecia a inmovilización do adxectivo e a imposibilidade de flexionar xénero e número, coma no seguinte exemplo extractado do CORGA etiquetado:

Fala claro, Bencho, ¿que queres dicir? [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Os casos que provocaron desconfianza e para os que a RAG propoñía (DRAG, 1997), e propón (DRAG, 2012), a adscrición á clase interxectiva son semellantes aos que reflectimos deseguido:

Só quedan unha trintena, os casos máis urxentes, e os pacientes, claro, protestan. [CORGAetq: GH2007-03-30/9]

Claro está, non lles corresponde aos redactores das Bases falar da estratexia que seguir en Galiza para un Estatuto de Nación ser aprobado no Parlamento de Galiza despois de ter o respaldo necesario nas eleccións deste ano 2005. [CORGAetq: ANT2005-03-03/5]

Claro que consideran aceptable a protección folclórica de "outras linguas", pero que metan o paternalismo por onde se enfían as agullas. [CORGAetq: ANT2008-07-10/68]

Xa que non temos un idioma común (agás o inglés, claro), a moeda é a icona (o ícono, segundo Boris) da cohesión europea. [CORGAetq: CG2000-11-22/80]

Tampouco nos satisfácea a solución achegada polo Xerais (2000 e 2009) de discriminar entre adverbio (=de maneira clara, claramente) e interxección (=evidentemente, certamente). Naturalmente, apreciamos a diferenza entre o uso adverbial pleno (*falar claro*) e o uso do aserto, máis ou menos enfático, (*Cres que lle gustará? Claro!*). Non obstante, cremos que esas diferenzas non afectan á categoría gramatical. Por unha banda, o seu comportamento encaixa co doutras locucións adverbiais de afirmación non cuestionadas: *por suposto, sen dúbida, dende logo...* que tanto poden aparecer como resposta única a unha pregunta, como facelo a xeito de inciso, como ir reforzadas pola conxunción *que* e o adverbio de afirmación ou negación, por exemplo. Pola outra, cremos que as diferenzas se deben ao seu distinto funcionamento na secuencia, claramente sintáctica no caso de *falar claro* e claramente discursiva en *claro está* e similares.

En conclusión, os casos dos que falamos non serán etiquetados no sistema de XIADA na clase das interxeccións, senón na dos adverbios.

4.4.7. Agrupacións con *outro*: ambigüidade no valor

A caracterización do atributo *valor –determinante, non determinante e determinante/non determinante–* para algúns casos da unidade *outro* obrigounos a revisar a súa entrada no lexicón e a engadir unha nota recordatoria no protocolo de desambiguación. Só ao enfrontarnos realmente coa etiquetaxe dos documentos fomos conscientes do problema, ata daquela inexistente na documentación de XIADA.

Por suposto, cando *outro* vai só ou acompañando a un substantivo a etiqueta que lle corresponde é clara: *non determinante* no primeiro caso e *determinante* no segundo. Por exemplo:

Pero unha cousa son os falantes e outra os coñecedores. [CORGAetq: GH2008-08-05/43]

"Non existe unha política familiar senón de servizos sociais, que é moi necesaria pero ten outra motivación, outros obxectivos e outra dotación orzamentaria", aseguran desde a FEFN. [CORGAetq: DLV2007-05-15/76]

E o mesmo sucede cando, aínda acompañado doutro determinante, forma grupo cun substantivo. Non hai dúbida respecto ao atributo *determinante* que lle corresponde no valor, tanto que sexa o primeiro determinante como o segundo. Por exemplo:

A xente antes das vacacións pode que visite algunha das mostras que baixo o paraugas sombrilla máis ben de Photoespaña están en diferentes sedes, pero de igual modo que as visitan en calquera outra época do ano e unicamente atendendo ao exposto nesas mostras particulares. [CORGAetq: ANT2008-07-10/45]

Entre estes dous extremos oín falar da existencia de moitas outras denominacións como son as vodas de coiro (aos dous anos), as de seda (aos 5) e as de estaño (aos 10). [CORGAetq: GH2008-08-05/95]

Pero avisen antes ó dentista de que xa non ando a navegar e mais de que non lle vou poder levar ningún outro loro. [CORGAetq: A. Conde, *A casa de Adara*, Xerais, 1996]

Logo de pechar o establecemento, ó quedar sós, retiráronse como tantas outras veces á trastenda. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

¿Quen pensa que, se non é mediante o autogoberno nacional, se poderá dispor dos propios recursos enerxéticos, ou reformar e modernizar a agricultura, ou explotar científicamente o litoral, ou diversificar a industria para facer fronte ás crises da construción naval e do automóvil, ou crear as estruturas de investigación, educativas e produtivas que permitan asimilar e desenvolver as tecnoloxías de vangarda, ou desenvolver en Galiza os ciclos produtivos e as actividades de transformación fundamentadas nas industrias e sectores básicos, ou potenciar a construción de vivendas, ou normalizar drásticamente a lingua galega e a propia cultura dentro dos sistemas educativo, informativo e económico, con todas as consecuencias industriais da cuestión, ou mellorar os transportes, ou aproveitar racionalmente a riqueza forestal, ou poñer os recursos e actividades financeiros ao noso servizo, ou dar traballo ás mozas e mozos que xa non emigran, ou mil outras cousas máis? [CORGAetq: MC1980-12/1]

O cambio do Goberno tivo como consecuencia, entre outras moitas cousas, un florecemento do xénero xornalístico da opinión. [CORGAetq: ANT2008-07-10/27]

A editorial francesa Hachette ven de adquirir o sector editorial de Vivendi ao completo, pasando así a controlar a produción do 50% do libro francés e a editorial española Anaya, propietaria da galega Xerais, entre outros moitos selos. [CORGAetq: ANT2002-10-31/57]

Quen considere que homes e mulleres están capacitados para desempeñar as mesmas funcións atopará serias dificultades para facer valer as súas ideas, pois se os dez mellores para rexer os destinos dunha comunidade son mulleres, veranse obrigadas a facer cinco ocos para outros tantos homes inútiles, que estarán alí de prestado, de floreiro. [CORGAetq: GH2007-03-17/81]

O director xeral da entidade financeira galega apuntou que, ademais deles, soamente outras dúas caixas "están presentes en toda a Península". [CORGAetq: CG2000-11-26/1]

Galicia e outras seis comunidades autónomas acordaron un documento para esixirlle ó Goberno a transferencia de competencias na creación de novas unidades xudiciais. [CORGAetq: GH2007-03-17/15]

Segundo o plan de viabilidade, reduciranse un total de 220 postos de traballo, a través de incapacidades por enfermidades e baixas incentivadas, concepto ó que se destinarán outros mil millóns de pesetas. [CORGAetq: CG1996-02-23/47]

Agora ben, cando *outro* se agrupa con numerais ou outros indefinidos e non hai un substantivo que determinar, xa non está nada claro cal debe ser o seu valor. Álvarez, Regueira e Monteagudo (1986) describen así esta peculiaridade:

Agrupado con numerais e indefinidos podemos consideralo núcleo da frase nominal ou adxacente, dependendo da súa situación na frase e do matiz que se queira dar: posposto ó resto da frase parece actuar como adxunto, pero anteposto pode desempeñar calquera das dúas funcións: *algún outro, ningún outro, moitos outros, calquera outro, tantos outros, tal outro*, frente a *outros moitos, outro calquera, outro(s) tanto(s), outro tal, outro(s) máis, outro(s) pouco(s), outros dous*; [Álvarez, Regueira e Monteagudo (1986: 243)]

Naturalmente, cando na cita se fala da posposición de *outro*, non se refiren os autores a que vaia tras un substantivo, xa que sempre se antepón a el (Álvarez e Xove, 2002: 509), senón a que nesa agrupación de determinativos *outro* ocupe a posición final. Nestes casos en XIADA, atribuímoslle o valor *non determinante*, valor que se corresponde para nós co "adxacente" da cita anterior. Por exemplo:

Se se impón a calidade, sen trampas, moito do que hoxe chaman ibérico pasará a ser porco de granxa, que aínda que sexa daquela raza terá unha carne como calquera outra, porque non terá os beneficios da alimentación con belotas de aciña e sobredo. [CORGAetq: CG2001-10-10/35]

Con todo, hai outros que se fosen xefes do mundo traballarían para lograr que os nenos con discapacidade tivesen as mesmas oportunidades que calquera outro. [CORGAetq: GH2008-08-05/31]

O pasado sábado, 26 de febreiro, o Foro do Novo Estatuto, integrado por relevantes intelectuais galegos como Xosé Luís Barreiro Rivas, Paulo González Mariñas ou Ramón Villares, entre

moitos outros, entregáronlle ao BNG un documento de conclusións que, en forma de bases, condensa e sintetiza varios meses de traballo e centos de achegas tanto de expertos como de cidadáns que, a través da páxina web do Foro, quixeron contribuír libremente a deseñar a Galiza do século XXI. [CORGAetq: ANT2005-03-03/20]

Ningún outro había coma el. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

O mesmo atributo, *non determinante*, asignámosllo ao numeral ou indefinido que, formando grupo con *outro*, ocupa a posición final:

Penso que Photoespaña, XI Festival Internacional de Fotografía y Artes Visuales, sempre foi unha maneira como outra calquera de concitar a atención mediática e do público, unha grande operación de merchandising artístico, ben promovida por unha empresa privada como é a editorial e galería La Fábrica (as institucións colaboradoras confían os catálogos destas mostras a La Fábrica) e unha escusa coma outra calquera coa que rematar o curso da programación artística madrileña. [CORGAetq: ANT2008-07-10/45]

Antonio Cajide Gulín é un colleiteiro que pertence a unha familia de longa tradición vitícola pero que se dedicou, coma outros moitos, durante unha parte da súa vida a actividades laborais diferentes que non tiñan nada que ver coa viña. [CORGAetq: ANT2008-07-10/63]

Adxudicouse á Sociedade Espazos Temolúdicos SA de capital maioritario de Caixa Galicia, a xestión privada dunha parte da obra, nomeadamente a explotación comercial e os locais de hostalaría co que se xerarán, en conxunto, case catrocentos empregos directos máis outros tantos indirectos, aspecto importante dada a elevada porcentaxe de paro existente na comarca coruñesa. [CORGAetq: ANT2005-03-03/21]

Repsol entra na guerra dos prezos e baixa entre 3 e 4 pesetas as gasolinas e outras dúas o gasóleo [CORGAetq: CG2000-12-21/43]

Desde 1991, cando Jesús Gil entrou a presidi-lo Concello de Marbella, o informe fiscalizador detecta unha nova política de constitución de sociedades municipais encargadas de xestionar servizos que antes prestaba a corporación, e en total constituíu once sociedades mercantís e adquiriu participacións noutras tres. [CORGAetq: CG1999-05-04/52]

De acordo tamén con Álvarez, Regueira e Monteagudo (1986: 243), cando *outro* ocupa a primeira posición nunha agrupación con outro indefinido ou un numeral sen que determinen a ningún substantivo, mantemos a ambigüidade do atributo *valor*, ao asignarlle o carácter *a*, que na posición que ocupa na cadea de etiquetaxe dun indefinido se le como *determinante/non determinante*. Así pois, ao *outro* dos exemplos anteriores correspóndelle respectivamente a seguinte análise:

outra calquera: IafS [Indefinido, determinante/non determinante, feminino, singular]

outros moitos: Iamp [Indefinido, determinante/non determinante, masculino, plural]

outros tantos: Iamp [Indefinido, determinante/non determinante, masculino, plural]

outras dúas: Iafp [Indefinido, determinante/non determinante, feminino, plural]

noutras tres: Iafp [Indefinido, determinante/non determinante, feminino, plural]

Agora ben, como debemos etiquetar o indefinido que ocupa a primeira posición nunha agrupación con *outro* cando non acompañan ningún substantivo?

Posiblemente sería necesario discriminar entre os casos en que *outro* pode ocupar calquera das posicións:

Antonio Cajide Gulín é un colleiteiro que pertence a unha familia de longa tradición vitícola pero que se dedicou, coma outros moitos, durante unha parte da súa vida a actividades laborais diferentes que non tiñan nada que ver coa viña. [CORGAetq: ANT2008-07-10/63]

A manifestación enmárcase nun movemento internacional de protesta pola guerra de Iraq, no que tamén participarán Barcelona, Sevilla e Zaragoza, aínda que, segundo as fontes convocantes, outras 15 cidades españolas desenvolverán concentracións de protesta contra a intervención española na guerra de Iraq e a presenza de tropas de Estados Unidos no país, así como noutras tantas europeas e americanas. [CORGAetq: GH2007-03-17/62]

Boa como calquera outra. [CORGAetq: ANT2005-03-03/71]

daqueles outros nos que só é posible na primeira, coma en

Desta forma, e preguntados neste termo, o 69,6% dos cativos galegos prioriza que o traballo lle guste, fronte ó 16,5% que antepón gañar moitos cartos e outro 13,9% que non se decide ó crer que ambas son igual de importantes. [CORGAetq: GH2008-08-05/29]

Ademais, un veciño de Vigo que responde ás iniciais A.S.A. foi imputado onte por distribuír e posuír imaxes de pornografía infantil a través de internet, dentro dunha operación contra a corrupción de menores que se desenvolveu en oito provincias españolas e supuxo a detención de nove persoas e a imputación doutras once. [CORGAetq: GH2007-03-17/36]

e finalmente, daqueles nos que só é posible en posición final, tales como:

Porque até nos momentos máis duros, ou tal vez neses máis ca en ningúns outros, temos dereito a ser donas das nosas vidas. [CORGAetq: ANT2005-03-03/67]

E é que os neurocientíficos Marc Hauser e Antonio Damasio acaban de contar en Nature que, se a unha persoa co córtex prefrontal ventromedial danado, como Gage, un lle pregunta se mataría o seu fillo para salvar dez, dez fillos dalgúns outros, a tal persoa diríalle que si, por pura matemática, dez a un, e por moi repugnante que poida parecer o cálculo, cando vostede e

mais eu salvaríamos o fillo, aínda a conta de condenar a dez e de mil malas noites, por moi repugnante que poida parecer o cálculo, neste caso tamén. [CORGAetq: GH2007-03-30/90]

Lamentablemente, non podemos afondar aquí nestas cuestións, mais aínda así estamos obrigados a realizar unha análise. As opcións que se nos presentan son dúas:

1) facer que sexa a posición na agrupación a que determine cal é o que actúa como *determinante*, de xeito que ese valor se lle atribuiría sempre ao primeiro indefinido, sexa *outro* ou sexa algún dos outros.

2) non especificar cal é o valor do indefinido que ocupa a primeira posición nunha agrupación con *outro*, deixando que persista a ambigüidade.

Ambas serían igualmente válidas e arbitrarias, pois non se orixinarían en argumentos de peso lingüísticos. Así e todo, a segunda das opcións ofrece varias vantaxes con respecto á primeira que nos levaron a considerala máis apta para XIADA. Entre elas destacan:

- A posibilidade de facilitar a recuperación nunha única consulta, por etiqueta, de todas as ocorrencias en que exista algunha agrupación con *outro*, independentemente da posición que este ocupe, cando non determinan un substantivo: Ia [Indefinido, determinante/non determinante].
- Concorde, así mesmo, esta segunda alternativa coa descrición para o *outro* que, en posición inicial en combinación con outro indefinido ou numeral non acompañados de substantivo, realizan Álvarez, Regueira e Monteagudo (1986: 243). A falta dun estudo polo miúdo destas agrupacións diversas fundamentado no funcionamento, significado e restricións de cada un dos elementos que conforman cada combinación concreta converte na opción máis neutra estender a non definición do atributo valor a todos os elementos que ocupan a posición inicial nunha destas combinatorias.
- Finalmente, mais non por iso factor menos decisivo, está o feito de tentar optimizar o rendemento do Etiquetador XIADA. A mesma etiqueta nas mesmas circunstancias incrementa a consistencia e coherencia do corpus de adestramento e facilita que o etiquetador poida inferir sen dificultade a etiqueta que corresponde.

4.4.8. *Propio*: indefinido ou adxectivo?

No VOLGa (2004), no DRAG (1997 e 2012), no Xerais (2000 e 2009) e no Ir Indo (2004) *propio* só aparece delimitado como adxectivo, fronte a *mesmo* para o que recoñecen nunha das súas acepcións a clasificación entre os indefinidos. Non obstante, nas *Normas* (RAG/ILG, 1995 e 2003) si se caracteriza como identificador, igual que na *Gramática* de Álvarez, Regueira e Monteagudo (1986) ou na de Álvarez e Xove (2002). En XIADA optamos por manter a distinción entre adxectivo e indefinido que propoñen Álvarez, Regueira e Monteagudo (1986: 251).

Deste xeito *propio* é etiquetado como indefinido

1) Cando, de forma similar a *mesmo*, precisa a identificación, intensificándoa. Por exemplo:

Inquedante paradoxo o dun país que importa as percepcións alleas do que lles pasa a outros para valorar a súa propia realidade. [CORGAetq: GH2008-08-05/91]

¿Esta falta de concreción non desactivará as bondades sociais da propia Lei? [CORGAetq: GI1996-06/30]

A supresión do IAE realimenta o proceso de destrución das estruturas do Estado de benestar facendo felices aos propios cidadáns que só verán os beneficios a curto prazo da diminución da súa carga impositiva aparente. [CORGAetq: ANT2002-10-17/30]

Segundo datos dos propios gobernos autonómicos, o presidente da Generalitat de Cataluña, Jordi Pujol, percibirá o próximo ano 14.942.850 pesetas, que se dividen en dúas partidas, unha asignación de 8.046.150 pesetas e 6.896.700 pesetas en concepto de representación. [CORGAetq: CG2000-11-27/2]

Cálculase que hai arredor de 15.000 escolas, e mesmo hai países (Australia, Alemaña e Suíza) nos que esta arte floral é estudada nas propias universidades. [CORGAetq: ANT2005-03-03/52]

Soñaba con que as propias rías deberan de se converter en privilexiados paraísos de produción pesqueira e marisqueira. [CORGAetq: CG2001-10-02/2]

Con isto, aseguran en Agrobyte, "as árbores representan máis diñeiro que o propio terreo antes da plantación". [CORGAetq: ANT2002-10-10/33]

O propio presidente, Emilio Botín, apuntou hai uns días na xunta xeral de accionistas que "as marxes financeiras tocaron fondo". [CORGAetq: CG1996-02-28/29]

2) Cando funcionando como núcleo e precedido de artigo equivale a *mesmo*.

Exemplos:

As mesmas fontes indicaron que durante un prazo de 45 días as dúas partes se comprometen, segundo o acordo asinado onte, a negociar un réxime de exclusiva, e polo tanto durante este período Air Europa non negociará a súa integración con outra compañía aérea nin Iberia fará o propio con outra compañía española. [CORGAetq: CG2000-11-03/74]

O documento elaborado polo Foro do Novo Estatuto ocupa xa, por si propio, un lugar sobranceiro na historia constitucional de Galiza. [CORGAetq: ANT2005-03-03/20]

Seguidamente, como decote, medio metín nas súas respectivas mangas o esquerdo, seguido do dereito, e por último fixen o propio co pequecho. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Víaos nun, todos xuntos, cunha carantoña torta e burleira, todos eles, mesmo pegados ó meu nariz, eles, os que paseaban a cotío polas nosas rúas, os habituais piropedores da dama, os propios, eles, todos, alí, en primeiro plano, a ollarme todos... [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Pola contra, caracterízase como adxectivo cando o seu valor é un dos seguintes:

1) Da propiedade daquel de quen se fale (sinónimo, *particular*; antónimo, *alleo*):

Moscoso, que nacera en México, fillo de emigrantes, e estudara medicina en Santiago de Compostela, integrouse con outros cantantes como Vicente Araguas, Xoan Rubia, Miro Casabella ou Suso Vaamonde na que entón se considerou como unha canción protesta baseada tanto en textos propios como noutros de Manuel María ou Celso Emilio Ferreiro. [CORGAetq: GH2007-03-30/39]

No referente aos ingresos, acentúase a dependencia estatal de xeito que as transferencias correntes representan o 74,47 por cento dos ingresos totais e as de capital o 8,87 por cento, habendo dúbidas razoables no grao de acerto nas previsións dos ingresos dos tributos propios e cedidos. [CORGAetq: GI1996-01/32]

Tamén defende a soberanía fiscal e tributaria, con capacidade para fixar os tributos e ter unha recadación propia, pois parte do principio que sen autonomía económica non hai autonomía política. [CORGAetq: ANT2005-03-03/3]

Blanco recoñeceu que Medio Rural "detectou" un "déficit de formación" tanto nas brigadas propias da Dirección Xeral de Montes, como no continxente de reforzo do verán, sobre o que admitiu que contaban cunha preparación "máis precaria". [CORGAetq: GH2007-03-17/1]

Martín Criado tivo unha posición relevante no sector inmobiliario, no que desenvolveu a promoción e a construción, a través de diversas sociedades inmobiliarias propias e compartidas, de máis de 20.000 vivendas. [CORGAetq: GH2007-03-17/1]

A expectativa de negocio non está tanto no xornal en sí, senón na concepción do que é o negocio da multimedia, da imaxe, o cable, a televisión..., onde sí ten que haber unha presenza

de entidades como a nosa, ben con capital propio ou ben financiando o proceso de desenvolvemento dun proxecto deste tipo. [CORGAetq: *GI1996-06/31*]

2) Que corresponde en exclusiva ou por natureza (=típico, característico):

Un dos tópicos que recaen sobre o sector pesqueiro en Europa é a consideración de que a pesca é unha actividade propia de países terceiromundistas. [CORGAetq: *GI1996-01/32*]

Heroes reais ou de ficción que expresan unha teoría do mundo e que neste caso ten bastante que ver coa visión da vida e da morte propia dos galegos (aínda que, por suposto, non só deles). [CORGAetq: *ANT2005-03-03/50*]

A pesar das dificultades que entraña o traballo da pedra e as belezas naturais que dela saen, os profesionais da cantería quéixanse do pouco retribuído que está este oficio, propio de artesáns con moita mañá e bo pulso. [CORGAetq: *ANT2008-07-10/65*]

Unha propaganda que comezou en 1945, e que vehiculizaron nomeadamente os norteamericanos, esforzouse en facernos crer que o de Hitler era un asunto extrahumano, propio dunha maldade irracional. [CORGAetq: *ANT2005-03-03/51*]

A proposta de novo Estatuto recolle o dereito a unha educación pública de calidade, inspirada nos principios do respecto á diversidade e da convivencia multirracial, da pluralidade política e relixiosa e nos valores propios da tradición cultural de Galiza e da historia común europea. [CORGAetq: *ANT2005-03-03/1*]

Trátase dun campionato centrado nos labores propios do campo e a muller da foto, Ainhoa Ansa, aparece en plena faena. [CORGAetq: *DLV2007-05-15/5*]

Cualidade favorecida polo atinado tratamento instrumental de todas e cada unha das composicións do disco, nas que se deixan translucir cadencias propias dos territorios evocados, como a "Morna sentida" a respecto de Cabo Verde. [CORGAetq: *ANT2008-07-10/42*]

3) Máis infrecuente ca os dous anteriores é o *propio* sinónimo de *axeitado*, de *verdadeiro* e *destacado*:

Certo é que a presenza da canción de autor latino e nomes propios como José Afonso, Serrat, Silvio Rodríguez ou Luís Eduardo Aute continúan a ser fontes claras para aprender e seguir camiñando, pero tamén é certo que estes artistas están falando das cousas que preocupan por aquí e que contan as súas historias na nosa lingua. [CORGAetq: *GH2007-03-30/112*]

Merce xa volve cunha bandexa, café (encol da bandexa), unha botella de güisqui (tamén na bandexa) e os vasos anchos propios do güisqui (escuso dicir que na bandexa). [CORGA: X. Cid Cabido, *Días contados*, Xerais, 1991]

4.4.9. Combinacións tipo *boca arriba*

No corpus de adestramento hai fundamentalmente dous casos nos que o valor *m* (*modificador*) do atributo *Valor* da categoría Adverbio o empregamos para dar conta non tanto da súa función no contexto no que se localiza como para relacionalo co substantivo ou sintagma inmediatamente anterior.

Un dos casos nos que sistematicamente se etiqueta como modificador responde ás construcións formadas por un substantivo e un adverbio de tipo locativo, como *boca arriba*. Basicamente os adverbios que entran nesta construción son *arriba*, *abaixo*, *adiante* e *adentro*¹⁸⁹. Vexamos algúns exemplos reais:

Quedou boca arriba, nunha posición inverosímil, coas pernas cruzadas e os brazos estendidos. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

E algunhas veces encamiñaba a besta monte arriba, polos camiños feitos por "eternidás de carros", e iamós parar á casa do Meleiro no Carballido. [CORGAetq: X. M. G. Trigo, *Extintos básicos*, Xerais, 1998]

o profesor puido, a duras penas, librarse das mans do agresor e conseguiu fuxir, laiándose escaleiras abaixo como se o matasen. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

O impacto do camiión provocou a ruptura do cilindro onde se almacenaba esta substancia pastosa, que se verteu e comezou a dirixirse augas abaixo nun tramo de catro quilómetros. [CORGAetq: GH2008-08-05/16]

A nós, xente de terra adentro, non nos amola o máis mínimo que falten cancións mariñeiras, de mariñeiros, e temas parecidos. [CORGAetq: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Eramos tres, ademais do dono, todos cantos formabamos o cortexo fúnebre, cemiterio de Boisaca adiante. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Fixo un xesto de despedida cunha man por riba do ombro, sen volverse, e continuou o seu cadencioso taconeo rúa adiante. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Álvarez e Xove (2002) non afondan neste tipo de construción, limitándose nada máis a constatar a súa existencia como modificadores de frase nominal; agora ben, malia a brevidade da referencia, introducen un aspecto interesante que os diferencia do resto de adverbios de lugar e que sería conveniente examinar máis polo miúdo:

¹⁸⁹ O ser forma non normativa non empece que a incorporemos neste grupiño, xa que, polo que a nós respecta, a súa presenza no CORGA nas combinacións que aquí tratamos así nolo esixe.

Como MOD nunha FN, os adv. de lugar e tempo precisan xeralmente da preposición *de*. Algúns adv. de lugar que teñen aglutinado un *a-* direccional (*abaixo, adiante...*) pódense usar posposto ó N sen intermediación dun nexo:

A roupa *de antes*, os rapaces *de agora*, os amigos *de aquí*, o veciño *de arriba*...

Polo río *abaixo*, pola rúa *adiante*, costa *arriba*... [Álvarez e Xove, 2002: 605]

Con estas combinacións exprésase ao mesmo tempo a dirección do movemento e o sitio por onde se realiza. Polas súas características parece que esteamos ante dous núcleos en contacto máis ca ante unha unidade multipalabra. Por unha banda, o substantivo non presenta variación xenérica nin é modificado por adxectivo:

Quedou boca arriba.

*Quedou *bocas* arriba.

*Quedou boca *xenerosa* arriba.

A existencia de substantivos plurais nestas combinacións non pode considerarse unha excepción, xa que se trata de substantivos que admiten a dupla variante (*escaleiras abaixo ~ escaleira abaixo*), sen que se produza en realidade unha oposición no número:

Ao velo no chan, corrín escaleiras abaixo e aniqueime ao seu carón. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Polo piso estendeuse unha regueirada de mexo que se precipitou en catarata escaleira abaixo. [CORGA: B. Iglesias, *A historia escríbese de noite*, Galaxia, 2001]

E xusto en canto terminou a súa sentenza ferinte coma unha puga oxidada, sen me dar lugar sequera a negar a miña condición de badulaque, escorreuse pola escaleira abaixo e caeu rebotando nos chanzos de granito coma un boneco desartellado. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Pola súa banda, o adverbio non admite a cuantificación de grao:

*Quedou boca *moi* arriba.

Os únicos casos, cremos, nos que o conxunto admite a modificación de grao son coa expresión *costa arriba*, mais aí xa non se expresa dirección nin lugar por onde se realiza a acción, senón que algo é moi dificultoso. Fronte ás combinacións que estamos vendo que se constrúen con verbos de movemento, neste último caso limitase a *poñerse* ou *facerse*:

O encontro púxoselles moi costa arriba ós locais tralo segundo gol do Atlético, nunha xogada persoal de Caminero que rompeu o fóra de xogo. [CORGA: *CG1996-02-04/84*]

A adolescencia é un tempo de moito medo, arroutada pero de moito medo, e cada vez que ía á casa de don Andrés ía con renovada curiosidade, pero ó mesmo tempo con medo a descubrir algún segredo que me fixese ver un mundo que eu non quería ver, que á mente daquel rapaz adolescente se lle facía moi costa arriba enfrontar. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Agora ben, fronte ao que se afirma para o castelán (Pavón Lucero, 1999: 610)¹⁹⁰ sobre que o substantivo non admite artigo nin outro tipo de unidade que poida actuar como determinante, en galego si é frecuente que apareza o artigo. Vémolos no último exemplo recollido, o de *escaleira abaixo*, e nos seguintes:

Total, fomos polo río abaixo. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Aínda podía escoita-lo balbordo da auga que fluía cara ó mar, polo cano abaixo, coincidindo co que facía a que entraba na cisterna, auga choída e fría. [CORGAetq: A. Conde, *A casa de Adara*, Xerais, 1996]

sucede, en troques, que queren botarnos a lingua de Cervantes pola gorxa abaixo, calcándolle co dedo do autoritarismo. [CORGAetq: *ANT2008-07-10/68*]

Só sentires como che escorre polas fazulas abaixo unha arroiada de lágrimas ben salgadas. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Por se acaso camiñou, lentamente e sen facer ruído ningún, pola habitación adiante ata se achegar ás roupas que xacían nunha butaquíña. [CORGAetq: A. Conde, *A casa de Adara*, Xerais, 1996]

Nin por todo o ouro do mundo emprendería unha carreira tola polo bosque adiante. [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

É certo, non obstante, que a aparición do artigo está condicionada, na nosa opinión, pola aparición inmediatamente antes da preposición *por*, que non fai máis ca incidir no lugar por onde transcorre a acción. É, pois, para nós unha mera variante expresiva.

Lamentablemente, aquí non podemos afondar máis nesta cuestión, pois senón desviaríámonos do obxectivo deste traballo, mais si queremos, coa análise que practicamos no corpus de adestramento, deixar constancia do comportamento especial destes adverbios en relación cun substantivo anteposto. Decisión que ademais non

¹⁹⁰ Unha vez máis, non atopamos estudos que tratasen este tema en galego.

empece, senón que facilita, a recuperación por parte do usuario final destes adverbios que en realidade non son núcleo da frase adverbial e que se atopan nunha relación case de interdependencia co substantivo que acompañan.

4.4.10. Combinacións tipo *anos despois*

O outro caso ao que nos referiamos máis arriba, para o que procedemos sistematicamente a realizar unha etiquetaxe non polo valor que lle corresponde a un adverbio nun contexto dado senón para facilitar a recuperación da información, responde principalmente aos adverbios *antes*, *despois* e *atrás*, caracterizados como modificadores no corpus de adestramento cando están modificados por unha frase nominal ou substantiva anteposta de carácter cuantificador¹⁹¹. Os exemplos que seguen son mostra desa etiquetaxe:

"Un ano antes de que comezasen a existir as duascenas millas de exclusión de cada país, España, a instancias de Víctor Moro, aprobou unha lei para a creación de empresas mixtas para a pesca. [CORGAetq: *TN1997-04/7*]

A tradición vitícola do Ribeiro vén do tempo dos romanos e séculos antes de que en Xerez houbera adegas xa Ribadavia era una especie de capital europea do viño. [CORGAetq: *CG2001-10-04/5*]

O líder de U2, Bono, reuniuse onte en Berlín co líder do Partido Social-demócrata Alemán, Kurt Beck, tres semanas antes da reunión do G8 en Heiligendamm. [CORGAetq: *DLV2007-05-15/42*]

Pouco tempo antes, un cargueiro francés partira en dous, perto das Cíes, un pesqueiro de Morás e foran moitos os mariñeiros de alí mortos no sinistro. [CORGAetq: *ANT2005-03-03/70*]

Un día iría a maquillarse para saír e, sen darse conta, como se xa o fixese miles de veces antes, Helena chamaríao por teléfono para lle pedir que fose cear á súa casa: [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Uns meses antes da inauguración, a Xunta Xeral de Accionistas acorda un cambio na localización da factoría baseándose nun informe técnico de Ciriaco Guisasola... [CORGAetq: *CG2000-12-03/77*]

Despedíronse tras un rito que se viña repetindo desde un ano atrás o mesmo día de cada mes. [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

¹⁹¹ O feito de que unha frase nominal de carácter cuantificativo poida desempeñar a función de modificador no interior dunha frase adverbial non é estraño. Entre outros autores recollen esta particularidade funcional Álvarez e Xove (2002: 219).

Estes datos veñen confirmar que a cultura da estabilidade está dando bos resultados, e as teses que apostaban anos atrás por unha política monetaria máis expansiva, co fin de crear emprego, reveláronse incorrectas. [CORGAetq: *CG1999-05-06/3*]

E a rapaza, cando sobe a cabeza de cabelo loiro e faciana enrubecida, olla para o único lugar onde pode mirar porque o seu esaxerado estrabismo diverxente obrígaa a dirixir o ollo dereito cara aquel home de grosas gafas e cara unha vella, dúas mesas atrás, sen elegancia á hora de comer churros, aínda que, en verdade, o único que tenta é ver o mar, en fronte, co ollo que pode, neste caso o esquerdo. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

Pero un ano despois a situación segue igual ou peor. [CORGAetq: *GH2007-03-17/38*]

E poucos días despois do éxito no torneo disputado en Girona, Xavi Román foi contratado polo equipo lugués, que rachaba así con esa carencia. [CORGAetq: *ANT2008-07-10/87*]

Quizais equivocadamente, cremos que a mellor maneira de etiquetar as frases adverbiais que levan como modificador unha frase nominal ou unha substantiva anteposta é marcar o adverbio como modificador, aínda que de feito sexa nuclear, pois á hora de realizar consultas e extraer os datos podemos diferenciar así estes casos dos do resto de frases adverbiais nas que un destes adverbios aparece illado ou só o modifica algún elemento posposto. Cremos que así damos conta igualmente da relación de dependencia que se establece entre a frase anteposta e o adverbio, aínda que en realidade a relación de dependencia que se produce sexa a contraria da que se sinala, feito que coa marcaxe do adverbio como nuclear non é posible, dado que non se entendería que existe relación entre eles. Finalmente, non cremos que esta etiquetaxe particular leve a equívoco, porque calquera usuario minimamente formado sospeitará que a marcaxe dese adverbio como modificador debe responder a algunha peculiaridade que o anotador pretende manifestar. Non obstante, a pesar de que non recibimos ningunha pregunta nin suxestión ao respecto¹⁹², é un dos aspectos de etiquetaxe pendente de revisión, por canto é dos poucos, senón o único, que non se corresponde no corpus de adestramento co papel que funcionalmente desempeña o elemento.

4.4.11. Enderezos web, números de teléfono, identificadores, etc.

Hai toda unha serie de elementos periféricos que non adoitan aparecer nunca tratados nas gramáticas dunha lingua, sexan descritivas ou prescritivas, recolléndose, como moito, algúns deles en manuais de estilo, mais sempre con someras indicacións de

¹⁹² No menú inicial da páxina web onde se sitúa o CORGA etiquetado incluíuse un ítem con esta finalidade.

como representalos e, loxicamente, sen entrar na súa delimitación ou clasificación categorial, labor que por outra banda non lles compete.

Con todo, calquera que coma nós teña que abordar a etiquetaxe completa das formas ortográficas dun texto ten que enfrontarse con este tipo elementos, asignalos, polo menos, a unha categoría gramatical, e decidir se se lles outorga algunha outra caracterización específica.

Son elementos, en xeral, que aparecen insertos nos diferentes enunciados ou que axudan a dispoñer e a organizar o discurso.

Do primeiro tipo, os máis habituais nos textos cos que tratamos son os enderezos web, enderezos de correo electrónico, publicacións dixitais, números de teléfono, matrículas de vehículos ou prefixos e sufixos illados que non acompañan un substantivo. Recollemos a continuación algún exemplo de cada:

Quen lle trouxo o suposto regalo a Z vendeuse ben como intermediario, pero seguramente o comprou na Obamastore (<http://store.barackobama.com>). [CORGAetq: GH2008-08-05/96]

Promoción do software libre desde Mancomun.org [CORGAetq: GH2007-03-30/49]

Ofrecen uns medios de contacto no teléfono [687 614 874](tel:687614874) ou no correo electrónico (manuelmaria@as-pg.com) [CORGAetq: ANT2005-03-03/37]

Podes consultar o cartel na web www.klubbers.com ou no teléfono [902499994](tel:902499994). [CORGAetq: GH2007-03-30/113]

Outras son máis novidosas e creáronse ao abeiro das novas tecnoloxías, como "todocolección.com" ou "vivescortada.com". [CORGAetq: ANT2002-10-17/28]

Cécilia Sarkozy, a muller do presidente electo Nicolas Sarkozy e "primeira dama" de Francia a partir do próximo mércores, non votou na segunda volta das eleccións presidenciais, segundo afirma o sitio de internet rue89.com. [CORGAetq: DLV2007-05-15/90]

Desde o portal tamén se pode acceder ás webs dos medios de comunicación do Grupo Correo Gallego, entre as que se contan www.galicia-hoxe.com, www.tierrasdesantiago.com, www.radiobradoiro.com _con emisión en directo da programación da emisora_ e www.correotv.es. [CORGAetq: GH2007-03-30/49]

Respecto do accidente rexistrado na Pobra, J.G.B.A., de 47 anos e veciño de Santander, faleceu ao chocar o turismo que conducía, un BMW matrícula [7147BRB](#), contra un camión Volvo _matrícula [4484CPT](#)_. [CORGAetq: GH2008-08-05/24]

Outra forma de incluír nunha palabra o significado de pé é utilizar no composto o prefixo ou sufixo grego pod (podómetro, podólogo, antípoda, monipodio, artrópodo, etc). [CORGAetq: ANT2005-03-03/27]

E esas deputadas... ista, ista, ista, ¿non habería que advertirlles da falacia? [CORGAetq: GH2007-03-17/81]

Do segundo tipo, encontrámonos fundamentalmente con identificadores –como a), A) ou 1._– que sinalan diferentes apartados ou aspectos que se queren realzar. Vexamos tamén algún exemplo:

a) Que as inversións teñan por obxecto específico algunhas das finalidades descritas no párrafo anterior. [CORGAetq: P1997-04/10]

B) Analfabetismo fiscal consistente en descoñecer a orixe e aplicación dos fondos públicos utilizados polo Estado. [CORGAetq: ANT2002-11-07/29]

2) ¿Atopas algunha incoherencia no texto? [CORGAetq: A. Álvarez Cáccamo, *Catapult*, Xerais, 1995]

2._ A cesión dunha porcentaxe, illadamente considerada, é inxusta e favorecedora das autonomías máis ricas. [CORGAetq: GI1996-06/41]

Dada a súa condición de elementos marxinais, consideramos pertinente clasificalos na clase Categoría periférica¹⁹³, onde tamén incluímos as siglas, os símbolos, as abreviaturas, etc., que aínda non aparecendo nas gramáticas si acostuman recollerse nos manuais de ortografía, por exemplo, no caso do galego, González Rei (2004: 263-299, 429-504, 547-562). Naturalmente, procedemos a discriminar dentro da Categoría periférica os distintos subtipos, denominando o que estamos tratando como *outros*. Cremos, non obstante, que para este tipo de unidades non ten obxecto ir máis alá da distinción realizada xa no subtipo, polo que non se considera pertinente ningún outro atributo. Así, todos os elementos que se remiten a este subgrupo, entre os que están os subliñados nos exemplos anteriores, levan a mesma etiqueta: Zo00 (*Categoría periférica, outros, xénero non aplica, número non aplica*).

4.4.12. Acurtamentos

Entendemos por acurtamento a supresión da sílaba final ou de varias que afecta especialmente a substantivos e adxectivos e que non fai variar o significado da unidade nin a súa delimitación categorial. É un fenómeno propio de rexistros coloquiais e do

¹⁹³ Véxase capítulo 1, apartado 1.3.2.18.

estrato xuvenil, consistente en acurtar vocábulos suprimíndolles xeralmente a(s) última(s) sílaba(s) e cuxo uso pode acabar por xeneralizarse ata o punto de recollerse xa uns cantos nas obras lexicográficas actuais.

Referímonos a unidades coma *tele*, *moto*, *pele*, *bus*, *bici*, *súper*, *progre*, *extra*, etc., presentes algunhas delas nos exemplos seguintes:

O mozo afirma que non tiña certeza de que falar polo móbil na bici infrinxise ningunha norma, pero ó comunicarllo á autoridade aceptouno como un mal menor. [CORGAetq: GH2007-03-17/40]

Adeus á viaxe ao Porto e adeus a vernos cada día, na súa casa ou na miña, para falarmos, beber, cear, ver algunha pele ou a tele, e mais o que despois lle dera por xurdir. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Os buses deixan atrás unha das maiores folgas que recorda o complexo: un seguimento de case o 100%. [CORGAetq: GH2007-03-30/9]

Mentres o progre refractario á televisión se deita pra madrugar e ler ben en tempo a cartilla (El País); encanto o somnolento fascista encama canda as galiñas pra erguerse cedo e estudar o catecismo (La Voz, El Correo), teño por norma trasnoitar en compañía de filmes canadianos (O declinio do imperio americano, As invasións bárbaras), das estrelas da NBA ou dun Ford Mustang descapotable, circulante por un documental de Canal de Historia. [CORGAetq: ANT2005-03-03/9]

A miña querida Silvia, a adorable Silvia, a perfecta Silvia, a progre Silvia, a enrollada Silvia, a equilibradísima Silvia, non entendo con que finalidade, contáralle a Xabier, a Fornos e naquel momento supuxen que xa a medio país, que eu tentara forzala pola noite, estando a pobre dela indefensa, durmida e eu borracho. [CORGA: X. Cid Cabido, *Días contados*, Xerais, 1991]

O abrigo, que colgo, mentres miro ao lonxe o título da película, consciente de que Xantia, que nunca se excitara cunha película porno, disfruta coa miña cara de miope, que, todo hai que dicilo, esaxero un pouco. [CORGA: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Dúas son as cuestións relacionadas coa etiquetaxe que lles afectan ás unidades acurtadas: i) a lematización e ii) a asignación dos atributos de xénero e número no caso dos adxectivos.

Polo que se refire á lematización, decidimos non reconstruír no lema da palabra truncada as sílabas que faltan con respecto á forma canónica, de xeito que o lema será a propia forma acurtada sen a desinencia do número plural, en caso de existir esta. Así, os

lemas das formas ortográficas subliñadas nas secuencias anteriores son, respectivamente, *bici, peli, tele, bus, progre e porno*.

Non nos parece adecuado fundamentar a adscrición do lema en función da aparición ou non da forma apocopada nun dicionario, pois é un criterio arbitrario que responde á difusión e asentamento de cada voz. Tampouco cremos axeitado recompoñer as sílabas suprimidas, dado que suporía neutralizar a diferenza existente entre os lemas de cadanseu rexistro lingüístico, un estándar e o outro coloquial. Non obstante, no momento en que se introduza no lexicón o campo do hiperlema, resolveremos a pexa da duplicidade orixinada por cambios de rexistro, xa que será posible relacionar os lemas anteriores, os apocopados, cos lemas completos ou non apocopados, ao habilitar a consulta por hiperlema (nos resultados terían que aparecer, por exemplo, todas as secuencias nas que concorresen os lemas *tele* e *televisión*¹⁹⁴) ou só por lema (ben os casos de *tele* ben os de *televisión*), segundo desexe o usuario.

Finalmente, no que concerne aos atributos de xénero e número das formas adxectivas acurtadas, actuamos igual ca con calquera outro adxectivo invariable: atribuímos o xénero e o número en función dos do substantivo que complementa. Así, nos seguintes exemplos *extra* caracterízase respectivamente como A0ms (*Adxectivo, grao non aplica, masculino, singular*), A0fs (*Adxectivo, grao non aplica, feminino, singular*) e A0mp (*Adxectivo, grao non aplica, masculino, plural*):

Segundo esta axencia, e a pesar dos desmentidos de Bankinter sobre unha hipotética ampliación de capital, esta operación podería ascender a uns 500 millóns de euros, e soluciona a necesidade de financiamento extra para mante-lo negocio actual de produtos hipotecarios. [CORGAetq: CG2002-05-02/10]

O prezo do gasóleo B pasou das 46,3 pesetas por litro en xaneiro do ano pasado ás 85 pesetas por litro na actualidade, o que significa que o Goberno conseguiu unha recadación extra vía IVA como consecuencia do incremento do prezo do gasóleo agrícola de 26.629 millóns de pesetas. [CORGAetq: CG2000-11-06/56]

O gasto dos consumidores representa dous tercios do Produto Interior Bruto (PIB) de Estados Unidos e en agosto creceu a penas un 0,2 por cento, a pesar dos cartos extra nos bolsos dos

¹⁹⁴ Pero non habería relación co lema *televisor* (s. m.) como equivocadamente indica o DRAG 2012 (dispoñible en http://www.realacademiagalega.org/diccionario/#loadNoun.do?current_page=1&id=44843 [consulta realizada o 2/7/2012]:

tele s.f. col. Forma abreviada de televisión (ac. 2, 3) e de televisor. *Pasou toda a tarde vendo a tele. Acende a tele.*

contribuíntes pola devolución de case 30.000 millóns de dólares dos tributos pagados o ano 2000. [CORGAetq: CG2001-10-02/43]

Hai ocasións, así e todo, nas que o adxectivo flexiona o número de forma parella a como pode facer un substantivo do seu mesmo tipo. Descoñecemos, non obstante, os factores que nuns casos fan aceptable a flexión e noutros a rexeitan. Compárense, por exemplo, nas seguintes secuencias os substantivos *progre* e *súper* e os adxectivos *extra* e *porno*:

Coma un herdo, porque se ben se pensa trátase dunha xeografía conquistada polos seus pais, os progres do 68, en tempos tópicos e utópicos. [CORGAetq: X. M. G. Trigo, *Extintos básicos*, Xerais, 1998]

Luíx fixo tempo observando asombrado aquel xeito vido a menos e fóra das modernas canles comerciais dos hiper e dos super ou das grandes áreas. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]¹⁹⁵

Tamén debido á estación, a roupa e o calzado que se usa é máis caro e provoca máis gasto, e o nadal obriga a unha serie de gastos extras, sobre todo en regalos. [CORGAetq: CG1996-02-16/1]

¿Por que non ven pelis porno? [CORGA: M. Rivas, *¿Que me queres, amor?*, Galaxia, 1996]

4.4.13. Soletreos e suspensións

Dende o punto e hora en que pretendemos dar cabida no corpus etiquetado absolutamente a calquera forma ortográfica que poida concorrer nun texto, sen andar eliminando as problemáticas, vémonos obrigados a tomar decisións de traballo sen dispoñer doutras referencias, é dicir, sen poder contrastar co que se fai noutros corpus con tales formas.

Nesta ocasión o obxecto da nosa descrición céntrase no tratamento en XIADA dos soletreos e as suspensións.

O primeiro deles, o soletreo, é claramente un fenómeno no que se mesturan aspectos fonolóxicos con morfolóxicos: a unidade segméntase nas sílabas ou letras que a constitúen. Pola súa banda, as suspensións omiten a parte final da palabra, que é sempre reconstruída mentalmente polo oínte. A diferenza das formas truncadas, as suspensións só son entendidas no contexto no que aparecen, sen que sexa posible

¹⁹⁵ Única mostra na versión 1.5 do CORGA de *súper* substantivo empregado en plural. [Consulta realizada o 3/7/2012]

eivalas a unidades comprensibles fóra dese contexto no que se producen. Por suposto, o emprego da suspensión na escrita débese a factores de índole estilística e non ten comparación posible con toda a casuística que pode orixinalas na fala. De todas formas, a nós non nos interesan os motivos que poden provocar suspensión ou soletreos máis ca na medida en que poden axudarnos a tomar unha decisión sobre como etiquetalos.

Cremos que as posibilidades de actuación se reducen a dúas: i) consideralos elementos periféricos, o que non nos gusta porque o grupo sintáctico no que están insertos é o característico dunha unidade central, e ii) tratalos coma calquera outro elemento dunha clase nuclear, opción pola que nos decantamos.

Así, tanto nos casos de soletreo como nos de suspensión, mantemos en *unidade* a unidade gráfica tal e como aparece escrita, cos guións normalmente separando as sílabas ou letras no primeiro e con puntos suspensivos finais no segundo. A etiqueta e lema serán os que lles corresponderían de non terse producido o fenómeno, ou sexa, os que lles serían outorgados se non aparecesen os trazos ou os puntos suspensivos fosen substituídos pola(s) sílaba(s) que faltan. Polo que respecta á unidade estrutural *forma*, hai un condicionante que nos impide tratar os dous fenómenos do mesmo xeito: a posible agrupación no segmento soletreado de máis dun elemento lingüístico, ou mesmo de secuencias enteiras. Vémolo mediante exemplos.

Os dous que seguen reproducen o tratamento que nos gustaría darlles no corpus de adestramento aos fenómenos de soletreo, o primeiro deles, e de suspensión, o segundo. Recollemos primeiramente a secuencia na que aparecen e, deseguido, amosamos representado en formato de só texto as etiquetas xml que darían conta da anotación desexada:

A miña cá-te-dra. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

```
<análise_unidade>
  <unidade>cá-te-dra</unidade>
  <constituínte>
    <forma>cá-te-dra</forma>
    <etiqueta>Scfs</etiqueta>
    <lema>cátedra</lema>
  </constituínte>
</análise_unidade>
```

O que sucede é que a vida dá moitas veces unhas hos..., unhas labazadas de moito mimo.
[CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

```
<análise_unidade>
<unidade>hos...</unidade>
<constituínte>
<forma>hos...</forma>
<etiqueta>Scfp</etiqueta>
<lema>hostia</lema>
</constituínte>
</análise_unidade>
```

O desiderátum conseguimos plasmalo na etiquetaxe do CORGAetq para as suspensións. Agora ben, nos soletreos só podería aplicarse este sistema cando afectase a unidades lingüísticas simples, cando a forma gráfica soletreada se correspondese cun único constituínte, pois como manter na *forma* os trazos se temos que desagregar en varios constituíntes? Non é posible porque se solapan dúas segmentacións que non se complementan, senón que se opoñen: a do soletreo e a dos constituíntes. Vémolo neste outro caso cunha forma verbal con pronome enclítico:

In-coá-ron-mo _mastigou con soberbia as sílabas_, un expediente disciplinario. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

```
<análise_unidade>
<unidade>In-coá-ron-mo</unidade>
<constituínte>
<forma>incoaron</forma>
<etiqueta>Vei30p</etiqueta>
<lema>incoar</lema>
</constituínte>
<constituínte>
<forma>me</forma>
<etiqueta>Rad1as</etiqueta>
<lema>me</lema>
</constituínte>
<constituínte>
<forma>o</forma>
<etiqueta>Raa3ms</etiqueta>
<lema>o</lema>
</constituínte>
</análise_unidade>
```

Así pois, nos casos de soletreo, dado que cremos fundamental achegar a mesma solución para as diversas manifestacións dun mesmo fenómeno, tan só daremos conta

na *unidade* da existencia de tal fenómeno, prescindindo na *forma* dos trazos que representan a silabación. Xa que logo, a análise que finalmente recibe *cá-te-dra* é a seguinte.

```
<análise_unidade>
  <unidade>cá-te-dra</unidade>
  <constituínte>
    <forma>cátedra</forma>
    <etiqueta>Scfs</etiqueta>
    <lema>cátedra</lema>
  </constituínte>
</análise_unidade>
```

4.4.14. Locucións

Calquera aspecto lingüístico no que poidamos pensar está case seguramente sendo obxecto de estudos continuos, ben porque hai lagoas nas súas descrições ben porque se aborda dende unha perspectiva distinta. Pois ben, o tema das locucións é un dos terreos máis pantanosos cos que nos podemos atopar, e só a necesidade de tratalas no corpus fíxo que nos achegaramos algo a elas. Por unha banda, apenas hai estudos específicos sobre as locucións que nós consideramos: prepositivas, adverbiais e conxuntivas. Pola outra, o seu elevado número e o amplo abano de posibilidades de formación e significación converten o seu estudo nun labor colosal. Nós, sen poder dedicarlles toda a atención que merecen, e tamén, por que non recoñecelo, sen posuír os coñecementos e interese que tal tarefa precisaría, procuramos sistematizar e ser coherentes no seu tratamento.

Deseguido realizaremos un resumo sobre o que, na nosa opinión, merece ser resaltado en cada un dos subtipos recoñecidos en XIADA, explicando antes cunha breve síntese a innecesariedade de catalogar no sistema as unidades discursivas.

4.4.14.1. Nivel oracional vs. nivel supraoracional

Nos últimos anos asistimos á proliferación nas distintas linguas de estudos que rompen o monopolio da oración como unidade maior de análise e ao nacemento do *texto* ou *discurso* como novas unidades que se converterán no centro de numerosas investigacións, dando lugar a novas disciplinas que acollerán estes estudos: pragmática, gramática do texto, análise do discurso, etc. Pois ben, neste panorama xorden novas

unidades, entre elas os *marcadores do discurso*, que acapararán a atención de moitos investigadores e darán lugar a múltiples páxinas.

O galego non é alleo a esta corrente, e así gozamos xa dunha *Gramática do texto* (Freixeiro, 2003) e de estudos parciais sobre algún dos conectores discursivos (por exemplo, Ferreiro, 2002 ou Domínguez Portela, 2006), sobre un tipo determinado (Freixeiro, 2005a, 2005b) ou sobre un rexistro concreto (Moreda, 2009).

É evidente que o enunciado, entendido como a unidade mínima de comunicación, pode estar representado por unha oración, pero tamén pode estar formado por moi diversas expresións que non teñen por que ser oracionais. Tamén cremos que, o feito de que se estableza unha relación entre dous enunciados non debe levar a considerar os elementos que serven de enlace entre estes como pertencentes a unha clase sintáctica, senón a unha clase discursiva.

A nosa atención en ningún momento se centra no *discurso* como unidade textual, senón na *oración*, e só dende unha perspectiva morfosintáctica. Énos indiferente se a conxunción *pero*, por exemplo, está actuando como conector discursivo ou como o elemento que permite contrastar nun enunciado dado unha Tese coa súa Antítese; o que si nos interesa é que, categorialmente, segue sendo unha conxunción.

Agora ben, hai ocasións nas que soamente se recorremos á perspectiva discursiva, a traballos específicos centrados na análise do discurso e nos diversos conectores, teremos axuda, malia que non sempre a necesaria, para delimitar categorialmente algunha das unidades que imos ver, ademais de explicar o diferente comportamento sintáctico e semántico de varias conxuncións e adverbios.

Por este motivo cremos fundamental aclarar, antes de continuar con esta epígrafe, que se entende por conector discursivo, que criterios se empregan para recoñecelos e que outros se utilizan para diferenciar entre eles conxuncións e adverbios.

No marco da hipótese teórica que concibe a comunicación humana como esencialmente inferencial e a existencia de unidades lingüísticas cuxo significado convencionalmente fixado na lingua condiciona o discurso en relación co contexto, xorde a seguinte definición para marcador discursivo:

Los 'marcadores del discurso' son unidades lingüísticas invariables, no ejercen una función sintáctica en el marco de la predicación oracional –son, pues, elementos marginales– y poseen

un cometido coincidente en el discurso: el de guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación. [Martín Zorraquino e Portolés, 1999: 4057]

Malia que entre os diversos especialistas non hai unanimidade sobre a denominación nin sobre o conxunto de unidades que integran o grupo, si hai uns trazos comúns nos que se fundamenta o recoñecemento e clasificación dos conectores discursivos.

Os principais criterios que permiten identificalos e distinguilos son, seguindo a Martín Zorraquino e Portolés (1999: 4059-4071), os seguintes:

1) Gramaticalización. Os marcadores do discurso son unidades lingüísticas invariables que perderon a capacidade de flexionar. Confróntese a) –conector– con b) –non conector–:

- a) Por se isto fora pouco, resulta que aínda non cobraron as expropiacións feitas o que, non obstante, non impide que teñan que seguir pagando a contribución por esas fincas. [CORGA: TE1977-34/1]
- b) Non obstante todo o anterior, coído que paga a pena facer o seguinte esforzo: [CORGA: ANT2006-12-14/5]

2) Posición sintáctica. As conxuncións preceden ao sintagma que introducen, mentres que os adverbios teñen en xeral grande mobilidade e poden incluso separar o Suxeito do seu Predicado ou os compoñentes do grupo do Predicado.

Si. Eso sábeno todos. Pero a cousa é moi vella. [CORGA: X. Fernández Ferreiro, *A morte de Frank González*, Edicións do Castro, 1975]

En fin, aqueles ventos recollen estas tempestades. [CORGA: CG1995-02-04/18]

O caso é que, ó mellor, tal vez, pero só quizais por suposto, pode que sexa el, se cadra, posiblemente ¿non?, un tanto non sabe, vaia, en fin, fetichista. [CORGA: CG1995-02-16/48]

Falarase, en fin, de Deus, que está no centro do islam e do cristianismo [CORGA: ANT2008-07-10/10]

Un non sabe, en fin, qué é mellor moralmente, pero si qué é normal, ou cordo, ou comprensible. [CORGA: GH2007-03-30/90]

É inmoral, en fin, que a mesma publicidade se teña feito necesidade. [CORGA: GH2007-03-01/68]

3) Marcas de entoación. Os adverbios conectores atópanse limitados como incisos pola entoación. Nunha pronuncia esmerada percíbese unha pausa posterior ao marcador e, ás veces, tamén unha posterior. Na escritura, esta entoación peculiar reflíctese habitualmente situando o marcador entre comas, inda que non é estraño que en ocasións non se faga.

4) Modificadores. Os adverbios discursivos, a diferenza dos que funcionan como complementos circunstanciais, carecen da posibilidade de recibir modificadores.

5) Negación. Os marcadores discursivos non poden ser negados.

6) Coordinación. Os marcadores discursivos non se coordinan entre si.

7) Foco e perífrase de relativo. Os conectores discursivos non poden nin focalizarse nin destacarse mediante perífrases de relativo.

Finalmente, os criterios que se empregan para diferenciar no interior dos conectores discursivos entre os membros da categoría Adverbio e os elementos da categoría Conxunción son os seguintes (Portolés, 2001: 51-56):

- Dúas conxuncións non vinculan sintacticamente as mesmas unidades, polo que non se documentarán as combinacións *e pero* ou *aínda que pero*. Non obstante, é habitual a coincidencia de *pero* e dun adverbio marcador no mesmo membro discursivo.
- As conxuncións sitúanse na posición inicial do seu membro discursivo sen estar delimitadas pola entoación, fronte aos adverbios que posúen unha maior mobilidade e un contorno entoativo propio.
- En estilo indirecto a conxunción *que* non pode preceder os marcadores que son conxuncións coordinantes, pero si os que son adverbios.

4.4.14.2. Locucións conxuntivas

4.4.14.2.1. Preposición/locución prepositiva + que, locución conxuntiva? Ata que, a pesar de que, para que, ó que...

Se lembramos a táboa de locucións conxuntivas dos mil elementos máis frecuentes da versión 2.4 do CORGAetq con *que* que vimos ao inicio da epígrafe anterior,

Orde	Elemento gramatical	Categoría
123	aínda que	Locución
144	mentres que	Locución
150	xa que	Locución

é probable que da simple lectura das locucións precedentes se bote en falta a mostra, sequera, doutro subtipo de locucións que segundo unha clasificación máis tradicional (RAG/ILG, 2003: 183-186) debía figurar nesa lista. Pois ben, por unha banda a táboa anterior só pretendía ser un exemplo visual das diverxencias entre os dous CORGA, o etiquetado e o sen etiquetar, provocadas pola etiquetaxe á hora de elaborar os índices de frecuencias; por outra banda, unicamente recollemos as unidades multipalabra conxuntivas das cales *que* é parte integrante mais non amalgamada (fronte a *porque* ou *anque* por exemplo); en terceiro lugar, seleccionamos as locucións con *que* que aparecen entre os primeiros mil elementos máis frecuentes da versión 2.4 do CORGA etiquetado e, finalmente, se non hai máis elementos nesa listaxe, só pode deberse a dúas razóns: unha, non aparecen entre os mil máis frecuentes, e/ou dúas, non son considerados locucións.

Sobre ambas as xustificacións, non necesariamente excluíntes, non nos achegan ningunha información as estatísticas; teríamos que ben acceder á lista de todos os elementos gramaticais ben descender ao corpus etiquetado, ás análises, para comprobar que unidades como *para que*, *desde/dende/des que*, *a que*, *ata/até que*, *antes de que*, *despois de que*, etc. non están clasificadas no sistema de XIADA como locucións conxuntivas.

A razón, ao noso modo de ver, é clara e, sobre todo, congruente co criterio adoptado para a discriminación das locucións, sexan estas adverbiais, prepositivas ou conxuntivas: a fixación e o significado diferenciado do da suma dos elementos que a compoñen.

Tradicionalmente tendeuse a considerar locución conxuntiva case calquera combinación de preposición ou adverbio seguida de *que*. Para demostrar tal aseveración basta con fixarnos nun subgrupo das conxuncións e locucións, por exemplo as concesivas, que achega a *Gramática galega* (Álvarez, Regueira e Monteagudo, 1986):

Outros nexos concesivos son *mesmo que, así, pese a que, a pesar de que, malia que, máis que, por máis que, por moito que, ben que, mal que, por* + adxectivo + *que*, etc. [Álvarez, Regueira e Monteagudo, 1986: 529-530]

ou a última edición das *Normas* (RAG/ILG, 2003: 184):

a pesar de que	malia que
a pouco que	mesmo que
aínda que	nin que
así	por (+adx.) que
ben que	por máis que
inda que	por moito que
máis que	por pouco que
mal que	

Esta consideración mantense en descrições lingüísticas xenéricas:

As locucións conxuntivas poden ter diversas formas, entre as que salientan as combinacións de adverbio ou locución adverbial + conxunción (*así coma, mal que, mesmo que, máis que...*), preposición ou locución preposicional + conxunción (*desde que, pese a que, agás se...*), participio + conxunción (*visto que, dado que, posto que...*), preposición + adverbio + conxunción (*a pouco que, por máis que, de aí que...*), substantivo + conxunción (*de xeito que*), etc. [Hermida, 2004: 203]

e noutras máis especializadas:

Mais son moitos os casos de relacionantes clausais que están formados por máis dunha palabra e que por conseguinte constitúen unha locución conxuntiva ou locución complexa (*aínda que, por causa de que, para que*), que pode constar de dous membros (*ou... ou, quer... quer, sexa... sexa, non só... mais tamén, de tal modo que*). O número de conxuncións simples é, na realidade, bastante reducido, entanto que as locucións conxuntivas se van incrementando por transferencia doutras categorías gramaticais, en moitos casos preposicións ou adverbios seguidos de *que* (*para que, sempre que, xa que, de xeito que*), á función propia das conxuncións. [Freixeiro, 2000: 579]

En cambio, Álvarez e Xove (2002) maniféstanse en contra desta tradición e propoñen non analizalas como locucións conxuntivas:

Agora ben, noutras funcións fórmanse sintagmas máis ou menos constantes en que interveñen adverbios e preposicións (ou locucións de ámbalas clases), que por veces son vistos como locucións conxuntivas con distintos graos de gramaticalización (*para que, desde que, a pesar de que, despois de que...*; nalgún caso mesmo soldada xa nunha única palabra gráfica: *porque, conque*). Con todo, cando nesos casos é posible unha oración amodotemporal sen *que*, semella

preferible analiza-las que aínda manteñen independencia formal como un sintagma integrado por prep. + conx. *que* + oración modotemporal ou por adv. (N da FAdv.) e prep. + conx. *que* + oración amodotemporal (Mod. do N da FAdv.); consonte isto, só poderían ser analizadas como locucións conxuntivas as que non permitisen o anterior:

Deulles leña a todos *para que* quentasen as casas [CC final: *para* + oración con *que*]. [Álvarez e Xove, 2002: 658]

Ao noso modo de ver, malia que concordamos por completo coa conclusión manifestada en Álvarez e Xove (2002: 658), a explicación é algo máis sinxela. No que se refire a preposición/locución prepositiva + *que*, dado que unha preposición pode ter como termo tanto unha frase nominal, substantiva, etc. como unha cláusula de infinitivo ou unha cláusula con *que*, é totalmente innecesario o cambio de categoría que supón reinterpretar preposición + *que* > locución conxuntiva:

Ademais para eles é unha vía de escape. [CORGAetq: ANT2005-03-03/72]

"Pero tamén significa un urbanismo pensado para os cidadáns e non para os especuladores. [CORGAetq: GH2007-03-30/23]

Neste caso a sala é pequeneira pero ampla, moi apta para actuacións musicais de esas que merecen ser escoitadas en silencio. CORGAetq: GH2007-03-30/112]

A medalla de prata foi para Alemaña e o bronce para Francia. [CORGAetq: GH2007-03-30/77]

Ademais, alertou sobre o risco de abusar do debate sobre as pensións para evitar "escenarios de pan electoral para hoxe e fame social para mañá". [CORGAetq: CG1999-05-14/45]

Lariyani sinalou que a "brusquidade do comportamento" de Londres leva a Irán a "optar pola vía xurídica" para abordar o asunto dos militares detidos. [CORGAetq: GH2007-03-30/61]

A incógnita segue a ser canto tempo se precisará para que os cambios sexan significativos. [CORGAetq: GH2007-03-30/61]

Evitamos, así mesmo, caer na incongruencia de tratar en categorías distintas os elementos introdutorios dunha cláusula de infinitivo (elemento prepositivo) e os dunha cláusula en forma persoal (elemento conxuntivo), ao considerar en ambos os casos preposición ou locución prepositiva; loxicamente, cando segue cláusula con verbo en forma persoal o elemento prepositivo vai seguido de conxunción subordinante:

Aumenta en 2.000 euros a dotación do Festival Internacional de Cine Euro-Arabe Amal, ata acadar os 37.000 euros. [CORGAetq: GH2008-08-05/42]

A Sociedade continúa vivindo uns anos máis ata que en 1908 Francisco Argüelles y Sierra embarga os bens da empresa". [CORGAetq: CG2000-12-03/77]

A pesar de recoñecer que agosto é "historicamente" o mes de máximo risco de lumes, recordou que a función da Administración é "perseguir os delincuentes" e "apagar" os focos que se rexistren. [CORGAetq: GH2008-08-05/5]

A pesar de que o propio presidente do executivo autonómico, Manuel Fraga, recoñeceu publicamente que os orzamentos "amosan luces e sombras", aínda non se solicitou unha revisión dos mesmos. [CORGAetq: ANT2002-10-03/31]

A moeda única pode tocar fondo ó chegar ós 0,8410 dólares e os analistas sitúan o seu teito nos 0,8520 dólares. [CORGAetq: CG2000-11-01/61]

A través dun acordo polo que achega á filial que ten no país veciño e desembolsa 10 millóns de euros, Indra adquire un 60 por cento da nova compañía, que se pasará a denominar Indra CPC, e mantén a opción de compra, ó que transcorran tres anos, do 40 por cento restante. [CORGAetq: CG2002-05-11/67]

4.4.14.2.2. Adverbio + de = locución prepositiva? + que, locución conxuntiva? Antes de (que), despois de (que), logo de (que)

Algo similar sucede coas, para algúns¹⁹⁶, locucións conxuntivas formadas a partir das supostas locucións prepositivas *antes de* e *despois de*, ás que, salvando as distancias, permitímonos engadir *logo de*, e mais a conxunción *que*. De seguido tentaremos xustificar a non consideración pola nosa parte das tres locucións prepositivas, para a continuación facer o propio cos seus correlatos conxuntivos.

En xeral, os adverbios base denotan, no tempo ou no espazo, prioridade o primeiro *–antes–* e posterioridade os outros dous *–despois, logo–*. Nos tres casos constitúen sempre o núcleo dunha frase adverbial, modificado en ocasións por unha frase preposicional encabezada con *de*. Así:

- poden non estar modificados:

PSA e a Zona Franca acabarán entendéndose, pero para iso tiveron que escenificar antes o seu desencontro nos medios. [CORGAetq: ANT2005-03-03/17]

Pero ningunha personalidade estadounidense en campaña, incluído Obama, se reúne con quen humillou a bandeira estadounidense e despois fixo saír ás súas tropas de Iraq nunha fuxida non negociada. [CORGAetq: GH2008-08-05/96]

¹⁹⁶ Por exemplo, RAG/ILG, 2003: 186.

Así, comparou estes votos cos "bonos lixo nos que moita xente investiu nos anos 80, que ofrecían beneficios xugosos e rápidos e logo resultou ser unha ruína". [CORGAetq: GH2007-03-17/21]

- poden estalo por unha frase preposicional cuxo termo non é unha cláusula:

Unha delas é que cando se rescata a concesión antes dese prazo, hai que indemnizar e iso permite obter recursos para un novo emprazamento. [CORGAetq: ANT2005-03-03/19]

Despois de varias probas realizadas en Madrid, no Canal de Experiencia Hidrodinámico do Pardo (CEHIPAR), que se saldaron con resultado positivo, a empresa ten previsto facer un novo control en mar aberto de cara a novembro deste ano. [CORGAetq: ANT2002-09-12/29]

Xa en breve, logo desta viaxe a Galicia, temos por diante en abril algunhas actuacións en Suíza e Alemaña, pero penso que imos facer este ano un descanso, unhas vacacións dende finais de ano. [CORGAetq: GH2007-03-30/111]

- poden estalo por unha frase preposicional cuxo termo é unha cláusula con *que*:

O pasado domingo, o presidente do club, Giorgio Pedraneschi, rompeu o trato logo de que os ingleses amosasen a súa inquietude sobre unha suposta lesión de xeonllo do colombiano e pedisen unha redución do traspaso, fixado en ó redor de 1.250 millóns de pesetas. [CORGAetq: CG1996-02-08/116]

O xogador albaceteño reclamou máxima atención para o encontro do próximo sábado ante o Deportivo da Coruña, despois de que a metade do persoal teña a mente posta nas súas seleccións durante toda a semana. [CORGAetq: GH2007-03-30/74]

Pouco antes de que enchesen a sala Capitol de Santiago de Compostela o sábado pasado, o vocalista deste combo chamado Skalariak, Juantxo, facía en Radio Obradoiro un repaso á saude desta formación a reivindicar. [CORGAetq: GGH2007-03-30/111]

- ou mesmo poden estalo por unha frase preposicional cuxo termo é unha cláusula de infinitivo:

Máis aínda, a primeira cuestión a dilucidar antes de avanzar afirmacións aventuradas é a de cifrar o fenómeno en discusión. [CORGAetq: GII996-06/34]

"despois de levar publicado en tantos formatos diversos, que priman o visual", tamén tiña unha certa ansiedade agora por publicar o que é un "libro normal", un libro para ler. [CORGAetq: GH2008-08-05/46]

Logo de analizar a de Vigo e A Coruña, presentou un estudo sobre as comarcas de Ferrol, Eume e Ortegal. [CORGAetq: ANT2008-07-10/70]

Non sabemos moi ben cales son as razóns que levaron a algúns autores¹⁹⁷ a considerar estes dous adverbios (*antes, despois*) seguidos da preposición *de* locucións prepositivas, pois nin posúen un significado diferenciado do da suma dos seus compoñentes, nin se comportan como preposicións, xa que a principal característica dos membros que integran esta clase morfolóxica é a de introducir o elemento que se denomina TERMO, do cal, salvo en contextos moi especiais¹⁹⁸, non pode prescindirse.

Así pois, desbotada a hipótese da súa consideración como locucións prepositivas, é evidente que, polos mesmos motivos que xa vimos para preposición + *que*, tampouco cando a *antes, despois* e *logo* lle segue *de que* son locucións conxuntivas, senón simples adverbios modificados por unha frase preposicional cuxo termo pode ser unha cláusula.

4.4.14.2.3. Antes que/ca, despois que/ca, logo que

Tanto *antes* coma *despois* poden considerarse adverbios comparativos sincréticos que inclúen na súa significación o comparativo *máis*, o que xustificaría a presenza ao seu lado do adverbio *moito* en lugar de *moi*: *moito antes* ou *moito despois* fronte a **moito lonxe* ou **moito amodo*. Evidencia, así mesmo, desta natureza comparativa é a aparición, como introdutor do segundo termo da comparación, da conxunción *ca*, xunto con *que* e coa preposición *de*:

Antes de min, houbo moitos narradores dispostos a aliviar a alma coa linguaxe; namorados cheos de señardade, de ausencias, e marfin, e pérolas, e adverbios de modo. [CORGA: X. Sierra Veloso, *Os ollos do rei de copas*, Edicións do Cumio, 1994]

_Outros antes ca ti abrigaron esas dúbidas, irmán _respondeu o xove frade_. [CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]

Todos tres preferiamos andar mar afora antes ca seguir vendo casas derrubadas e xentes chorosas. [CORGA: X. Neira Vilas, *Galegos no Golfo de México*, Edicións do Castro, 1980]

¹⁹⁷ Por exemplo, Álvarez, Regueira e Monteagudo (1986) ou RAG/ILG (1995 e 2003).

¹⁹⁸ *Por* e *para* permiten a omisión do seu termo en conversas sempre que a entoación substitúa o pronome interrogativo. No CORGA só atopamos exemplos con *por*; algúns deles son os seguintes:

[CORGA: C. G. Reigosa, *Narcos*, Xerais, 2001]
_Fun ata a casa de Belarmino Castaño.

_¿Por?

_Seguía sendo o máis sospeitoso, para min.

[CORGA: A. Riveiro Coello, *As Rulas de Bakunin*, Galaxia, 2001]

_Tiveches sorte co compañeiro de cela.

_¿Por?

_Bernardo Figueiras é toda unha personalidade no cárcere.

Os nenos bilingües constrúen antes que os monolingües frases "máis longas e complexas" sendo o seu vocabulario semellante, segundo destaca un estudo realizado por investigadores do Departamento de Psicoloxía Evolutiva e de Educación da Universidade de Santiago (USC). [CORGA: GH2007-03-29/45]

As cifras de MSN Serch Internacional apuntan a que un 71 por cento dos casos busca respostas rápidas na web antes que preguntarlle a ninguén ou consulta-lo en libros ou periódicos. [CORGA: CC2006-06/28]

Quizais os casos máis difíciles de refutar como locucións conxuntivas son aqueles de *antes que e despois que* en que non se manifesta tanto a preferencia como a anterioridade ou posterioridade con respecto a unha referencia temporal:

Esta notación procedía do imperio bizantino, pero antes que Occidente a adoptase, existía outra forma de representar as notas que consistía nunha serie de ganchos, vírgulas e puntos (signos derivados dos acentos das palabras) colocados sobre textos. [CORGA: M. Groba Lorenzo, *Introdución á linguaxe musical*, Galaxia, 2006]

O de Gándara, despois que entregaron os carros, demorouse entre os grupos, xentes e caras que endexamais denantes mirara pero que o saudaban agora, sen coñecelo, pero coma se todos fosen vellos camaradas felices de volveren saber uns dos outros logo de moito tempo. [CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

Non obstante, malia a problemática que suscita a clasificación destas dúas unidades, preferimos non consideralas locucións conxuntivas en ningún caso e si tratalas como adverbios seguidos dunha cláusula con *que*, dado que por regra xeral as conxuncións non admiten cuantificadores, ben adverbiais ben nominais, mentres que *antes que e despois que* si o fan:

Se as garantías definen a democracia, no Estado de Israel non existen e moito antes que Ariel Sharon sacase a escavadora a pasear, en Gaza, Xerusalem e a Ribeira Oeste derrubaban casais palestinos a diario, cegaban pozos, torturaban e eivaban detidos, moitos deles nenos. [CORGA: ANT2002-09-26/43]

Cando me atacou a cabra, pouco despois que baixase do taxi que me achegou a aquela mansión, por un instinto de protección, cometín o erro máis incalificable que ninguén podería imaxinar nun namorado da pintura: [CORGA: A. Álvarez Cáccamo, *Xente de mala morte*, Galaxia, 1993]

Desde o pasado 3 de novembro, un día despois que un coche bomba estoupou nos xardíns de Clara Campoamor de Barcelona, o teléfono 900 100 091 recibiu unhas 200 chamadas diarias. [CORGA: CG2000-12-22/69]

Pola contra, o adverbio *logo* non ten implícito no seu significado a comparación nin *logo que* admite ningún tipo de cuantificador, razón pola que si a consideramos locución conxuntiva subordinante no sistema XIADA:

Castro Páramo formulou esta invitación, de "tirón de orellas", logo que Núñez Feijoo instase os seus a que deixen gobernar nas municipais á lista máis votada. [CORGA: GH2007-03-29/17]

4.4.14.2.4. Sempre que

Existen outras combinacións de adverbio mais conxunción *que* as cales, dependendo do seu comportamento semántico-sintáctico, reciben un tratamento dispar no noso sistema. É o caso, por exemplo, de *sempre que*. Evidentemente, non queremos referirnos ás combinacións casuais de adverbio *sempre* seguido de *que* conxunción subordinante (exemplo 1) ou de *que* relativo (enunciado 2), senón a aquelas ocorrencias en que ambos os elementos conforman unha unidade (casos 3, 4, 5 e 6):

- 1) dicíame sempre que se todo lle saía ben non habían tardar en o facer alcalde dunha vila importante ou mesmo gobernador, e eu marcharía canda el e lonxe de aquí, había ser a máis bonita da provincia, con roupas aínda mellores ca as que me el daba de cando en vez, con xoias de nácara e ouro, e teríamos fillos, catro, polo menos, e serían educados e habían facer carreira. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]
- 2) Segundo Quintana, "nestas eleccións, a escolla nunca estivo tan clara: ou o nacionalismo resolutivo, que se ocupa dos problemas da xente, ou os políticos de sempre que se ocupan de cousas que só lles importan a eles, como a unidade de España", con clara referencia ó PP. [CORGAetq: GH2007-03-17/20]
- 3) Como ocorría sempre que falabamos por teléfono, só cara ao momento da despedida me despechou o motivo da súa chamada. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]
- 4) Sempre que o corpo e a alma llo permitían, Laureano achegábase ó lugar dos feitos aniñando tal vez a esperanza de que un raio xupiterino puxera fin a semellante barbarie. [CORGAetq: X. M. G. Trigo, *Extintos básicos*, Xerais, 1998]
- 5) "Inicia-la UEM sen a participación de certos grandes países non é ideal, pero non haberá ningún perigo para a cohesión da Unión sempre que quede claro que entrarán no futuro e que a maioría o vaia facer nun prazo curto", di o documento. [CORGAetq: CG1996-02-07/103]
- 6) Non obstante, o gobernador do Banco de España destacou que os países que lograron ter uns orzamentos equilibrados, como o caso de España, teñen moita máis marxe para actuaren nestas situacións difíciles, sempre que non se poña en perigo o saneamento das facendas públicas. [CORGAetq: CG2001-10-09/32]

A clasificación e descrición de *sempre que* non está exenta de polémica. Así, Álvarez, Regueira e Monteagudo (1986: 537) recoñécena como nexo condicional, mentres que Freixeiro (2000: 586-588) e Hermida (2004: 215-218) só lle recoñecen o valor temporal. Álvarez e Xove (2002: 659-660), pola súa parte, resultan un pouco máis confusos, pois, se por unha banda falan de que adverbio mais *que* poden chegar a formar locucións conxuntivas e citan, entre outras, *sempre que*, deseguido exemplifican o contrario para *sempre que*:

A oración de *que*, precedida de preposición, pode ser modificador dun adverbio nos sintagmas adv. + prep. + *que* + verbo modotemporal: *antes de que*, *despois de que*, etc. Cando a oración de *que* segue inmediatamente ó adverbio, pode ocorrer que cheguen a formar locucións conxuntivas (*antes que*, *logo que*, *sempre que*, *así que...*), con distintos graos de gramaticalización. Doutra banda hai construcións en que a presenza de determinados adverbios (*aínda*, *mesmo...*) nun dos termos induce a establecer unha relación semántica de certo tipo entre as dúas oracións; se a oración é amodotemporal (de Xer.) abonda coa forma adverbial, coma nas frases, pero se é modotemporal cómpre a presenza de *que*, de onde se deduce que adv. e conx. non integran unha unidade conxuntiva, senón que é o adv. o que incide sobre a unidade que segue:

Aínda que veña Román non somos abondos [CC concesión: adv. *aínda* + conx. *que* e verbo modotemporal; cfr. "*aínda* vindo Román..."].

Non o vai facer, *mesmo que* llo pidas de xeonllos [CC concesión: adv. *mesmo* + conx. *que* e verbo modotemporal; cfr. "*mesmo* pedíndollo..."].

Eu fágovolo, *sempre que* vós poñáde-los materiais [CC condición: adv. *sempre* + conx. *que* e verbo modotemporal; cfr. "*sempre* poñendo..."].

chegando tamén a indicar que (cfr. 157-158):

O parentesco das oracións condicionais coas temporais e coas causais explica o uso de nexos característicos destoutras clases (*cando*, *sempre que*, *como*, *dado que...*), que dan lugar a verdadeiras oracións condicionais co verbo en Subx.:

Mariña comprométese a facelo, (*sempre e*) *cando* non teña que poñer por riba cartos do seu peto [co significado condicional 'en caso de que', non co temporal de 'no momento futuro en que']. Dixo que se encargaba do papeleo *sempre que non tivese que facer ela as colas*. (...)

Malia que no corpus etiquetado non atopamos aínda ningunha ocorrencia de *sempre que* con subxuntivo onde prime o sentido temporal, si se documentan no CORGA, como demostran os dous exemplos seguintes, polo que non podemos aceptar a

hipótese de que o modo verbal decide se o valor da secuencia en cuestión é temporal ou condicional:

7) Preguntada esta muller se a fumación con herbas máxicas lle subtraera as dores, esta contestou afirmativamente e aínda engadiu que cos remedios da acusada, unha muller podería ter un fillo sempre que quixer, alegre e contenta de velo, sen outra preocupación que a de criar o neno. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

8) A idea orixinal era que miña nai se mudara comigo para unha vivenda nova, nunha zona acabada de urbanizar, meu pai seguiría a vivir na nosa vella casa e eu podería visitalo sempre que quixera. [CORGA: M. Regueira, *L'affiche rouge*, Xerais, 2007]

Así pois, descartados 1) e 2) como ocasionais, atopámonos con que a combinación *sempre que* ten sentido temporal construída con indicativo (3, 4), mais tamén con subxuntivo (7 e 8), secuencias todas elas nas que a podemos conmutar por *cada vez que*, mentres que só a construción en subxuntivo pode dar lugar á interpretación condicional, conmutable por *se*, (secuencias 5 e 6).

O que nos falta por dilucidar é se esa diferenza de significado se plasma nunha diferenza categorial, é dicir, se ambas son locucións, se non o é ningunha ou se só unha delas o é. Dado que nin a tradición galega nin a portuguesa¹⁹⁹ nos axudan moito, botaremos man da española, pois a dita combinación tamén ten lugar nesa lingua e coas mesmas características. Non nos sorprende en absoluto que na *Nueva gramática de la lengua española* (RAE, 2009) se recoñeza a existencia de polémica en torno a esta expresión, mais si que sexan tan claros ao afirmar que hoxe en día téndese a considerar locución conxuntiva cando presenta valor condicional e adverbio seguido de pronome relativo cando o seu valor é só temporal:

La locución conjuntiva condicional *siempre que* alterna en ciertos registros con *siempre y cuando* (§ 47.10d, e), que carece, en cambio, de usos temporales: *Te lo diré {siempre que ~ siempre y cuando} prometas guardar el secreto. (...)*

Es más polémico el análisis sintáctico de *siempre que* cuando esta expresión tiene sentido temporal. Aunque algunos autores entienden que sigue siendo conjunción en estos usos, otros sostienen -parece que con mayor fundamento- que *siempre* es un adverbio cuantificativo en esta interpretación (equivalente aproximadamente a 'todas las veces') y que la partícula *que* que lo sigue encabeza en este uso unha subordinada relativa. Está a favor de este análisis el que la

¹⁹⁹ Nin Cunha e Cintra (2005: 582-583) nin Vilela (1999: 260) nin Mira Mateus et al. (2003: 705-710) inclúen nos conectores condicionais *sempre que*, inda que si nos temporais.

construcción pueda usarse para contestar preguntas encabezadas por *cuándo*, el que el adverbio *siempre* pueda negarse (*No siempre que te llamo me contestas*) y el que admita el modificador *casi* (*Casi siempre que salimos de paseo, llueve*). [RAE, 2009: 2475-2476]

Os argumentos que se empregan para descartar *sempre que* como locución conxuntiva no seu valor temporal e aceptala no condicional convéncennos por completo, pois vemos que as diferenzas se reflicten tal cal no galego: negación de *sempre*, modificación con *case*...

- 9) Tamén, era verdade, posuíra non poucas mulleres, unhas co seu propio consentimento, outras á forza, pero facíao cando as ocasións eran propicias, e non sempre que quería, pois eu non era dono máis que a medias de min mesmo. [CORGA: X. Costa Clavell, *O conde de Viloíde*, Sotelo Blanco, 1985]
- 10) Mais non sempre que unha persoa teña dificultades para respirar hai que pensar que é asmática. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]
- 11) Porque case sempre que un comeza a súa andadura literaria, normalmente, encamiñase cara á poesía. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]
- 12) Barruzaba aquela tarde e el andaba só, como lle pasaba case sempre que remataba o diñeiro. [CORGA: P. Martín, *A bisneta lercha. Contos de sempre para ler hoxe*, Galaxia, 2006]

Así pois, no sistema de XIADA soamente será tratada como locución *sempre que* cando o sentido condicional sexa inequívoco (secuencias 5 e 6), mentres que a combinación con valor temporal será desagregada en adverbio *sempre* e relativo *que*²⁰⁰ (exemplos 3, 4, 7, 8, 9, 10, 11 e 12).

4.4.14.2.5. Agora que, agora ben, ora que, ora ben

A expresión *agora que* adoita aparecer no noso corpus con valor temporal, en cuxo caso caracterizamos un adverbio nuclear seguido dunha cláusula de relativo cuxo pronome, clasificado morfoloxicamente en XIADA como Tnaa (*Relativo, non determinante, masculino/feminino, singular/plural*) desempeñaría a función de Complemento Circunstancial na cláusula na que se insire. Son mostra desta análise os seguintes enunciados:

¿Non soou xa a hora da Galiza europea agora que acaba de dar a hora do euro? [CORGAetq: TN1998-05/18]

²⁰⁰ Dado que no adverbio non aplican os atributos de xénero e número, sempre que o antecedente dun relativo é un adverbio, este caracterizarase como *Relativo, non determinante, masculino/feminino, singular/plural*: Tnaa.

Para esta medida, o BCE conta cunha marxe de manobra maior agora que minguaron as presións inflacionistas nos doce países do euro. [CORGAetq: CG2001-10-26/33]

Agora que parece que chegan os primeiros síntomas do cambio climático, a fatiga do planeta, as especies esquilgadas, os mares mortos, etcétera, sorprenden estas cousas. [CORGAetq: GH2007-03-17/85]

Nós defendemos a tese de que, como xa mantivemos para *sempre que* con valor temporal, dende o punto e hora en que o adverbio *agora* na combinación con *que* pode aparecer modificado, por exemplo por un adverbio en *-mente* ou *nin*, non pode constituír unha locución conxuntiva:

Nestes fríos e duros días de inverno, precisamente agora que a claridade diurna xa se deixa notar, os santos e os feitos cristiáns que celebra o calendario fan fincapé no triunfo da luz sobre as tebras e desta maneira, os doutores da Igrexa, aproveitan a similitude co tempo meteorolóxico para expoñer o triunfo da fe sobre as crenzas precristiáns e o paganismo. [CORGA: ANT2003-01-23]

É de supoñer que esta tendencia á baixa continuará irremisiblemente, especialmente agora que se poden sacar fotos co teléfono móbil a unha calidade de resolución de até dous megapixel. [CORGA: CC2006-01]

¿Por que agora, precisamente agora que tiña moza e traballo, proxectos e presente? [CORGA: X. Queipo, *Papaventos*, Xerais, 2001]

Pero iso non conseguía tranquilizala, pensando se cadra que fora capaz de cumprir igualmente a ameaza de tantas veces. Agora precisamente que ela me necesitaba, cando máis lle podía doer. [CORGA: X. A. Neira Cruz, *As cousas claras*, Xerais, 2001]

¡Nin agora que estás morta me valoras o que merezo! [CORGA: M. Seixas, *A velocidade do frío*, Xerais, 1996]

El non mo dixo. Nin sequera cando intentou bicarme e meterme man. Nin agora que estamos mellor. [CORGA: M. Vázquez Freire, *Anxos en tempos de chuvia*, Xerais, 1997]

Pradera queda reducido antes do previsto á mínima expresión (e menos mal que lle deixan algo), xusto agora que empezaba a dar un xiro interesante cara á actualidade. [CORGA: CG2001-10-10]

Corroborar, así mesmo, esta análise de adverbio mais relativo o feito de que os relativos, cando o seu antecedente é temporal, locativo ou modal, acostuman non aparecer precedidos da preceptiva preposición²⁰¹:

²⁰¹ Véxase por exemplo Hermida (2004: 155-156).

A fuga máis longa dos últimos tempos, que estivo salpicada con moreas de suposicións e conxecturas, non se saldou do xeito que agardaba todo o pobo español. [CORGA: *VG1995-03-04*]

Por último, asegura estar disposto a debater con Aznar "no sitio que queira". [CORGA: *CG2000-12-12*]

Que me chame á hora que sexa. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

E tamén a circunstancia de que cando se coordinan dúas cláusulas o adverbio non ten por que aparecer na segunda, malia que poida facelo:

Agora que o teño tan preto e que toda inseguridade, toda incertidume, desapareceron en min, todas menos unha, menos a derradeira, teño a certeza absoluta de non querer ser el. [CORGA: X. Miranda, *Historia dun paraugas azul*, Xerais, 1991]

Segundo se dicía, era mesmo posible que importantes equipos extranxeiros estivesen negociando o traspaso do xogador, agora que aparecía completamente recuperado das súas lesións e que os últimos acontecementos lle deran unha certa notoriedade, mesmo facéndoo aparecer como vítima da guerra no seu país. [CORGA: M. Forcadela, *Fóra de xogo*, Edicións do Cumio, 1993]

"Esto si que foi un bo golpe, _engadiu_ e xusto agora que estabamos descabezando a HB e que, por primeira vez, xurdiran grupos discrepantes coa liña oficial, chega esta lei do PP, que provocará que volvan recuperarse, e esto a nosoutros desfainos". [CORGA: *CG2002-05-04/7*]

Algúns estudosos, sobre todo no ámbito castelán por ser este un terreo moito máis traballado ca no galego²⁰², consideran que a combinación do adverbio *agora* mais *que* é unha locución cando encabeza unha construción na que equivale a “a propósito” ou “por certo”, normalmente parentética, onde o sentido temporal está perdido ou moi atenuado e se introduce na intervención un dato novo provocado por unha circunstancia conxuntural (alguén menciona algo que o fai recordar ou ao falante lle vén á mente por ver ou lembrar algo). Referímonos, como non, a casos coma os seguintes:

Agora que o dis, ¿quen vai conducir? [CORGA: X. Cid Cabido, *Días contados*, Xerais, 1991]

Muller: Aínda que, agora que o penso, antes non che deu ningún. [CORGA: R. Vidal Bolaño, *Bailadela da morte ditosa (Sete baileretadas de amor e unha de morte)*, Sotelo Blanco, 1992]

Agora que me lembro, non crean que a marabunta de mesocracia borracha que pulula pola zona do Orzán son parto dalgún xenio do marketing. [CORGA: *GI1996-07/25*]

Nós non cremos que esteamos ante unha verdadeira locución pois

²⁰² Por exemplo María Moliner (2007) ou Pavón Lucero (1999).

a) Por unha banda, admite construírse con diferentes verbos (*ver, dicir, lembrar, falar, pensar, etc.*) e estes aparecen cos argumentos que esixe o seu esquema sintáctico (por exemplo, *dicir* con Complemento Directo ou *falar e lembrar*, con Complemento de Réxime), co que se demostra a variabilidade da construción:

E agora que me lembro de "circunferencias"... ¿viron vostedes o outro día a manifestación das nove, orgaizada pola "Unión de Mulleres de non sei que", que portaban unha cartelada na que puxeran: Esiximos orgasmos libres e gratuitos pola Seguridade Social? [CORGA: L. Iglesias de Souza, *As alaudas do mencer. Pintura negra, en forma de folletín noveleiro, policiaco-humorístico-político-social e satírico, (tamén) en catro romances de cego, catro partes e certos comedios*, Edicións do Castro, 1989]

Agora que o dis, parece que non respira. [CORGA: M^a. X. Queizán, *Ten o seu punto a fresca rosa*, Xerais, 2000]

Entrevistador: Agora que falas de Sudamérica, alá estiveron os SFDK e fliparon ... [CORGA: GH2005-02-11/80]

b) Por outra banda, a suposta locución conxuntiva admite modificadores e mesmo ir posposta a unha locución adverbial discursiva digresiva:

Precisamente agora que falas de obras... elas teñen moito que ver coa túa visita. [CORGA: A. Rodríguez Baixeras, *O rei dos ameneiros*, Xerais, 1987]

A propósito, agora que o dis, ¿como saben os amigos do teu pai se teño móbil ou non? [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]

c) en terceiro lugar, malia que non o documentamos no CORGA, cremos que é perfectamente gramatical a coordinación sen ter que repetir o adverbio, o que dá mostras da súa falta de cohesión:

Agora que te vexo e que me acorda....

d) Finalmente, non sempre ocupa a primeira posición nin pode dicirse que sempre sexa marxinal, a non ser que se faga pensando na diferenza de todos coñecida entre actantes e circunstanciais (Freixeiro, 2000 ou García-Miguel, 1995):

Nunca o pensara, pero agora que mo dis... [CORGA: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

Entrevistador: _Vemos na vosa web, agora que nos falas do tema, que sodes socios oficiais de Dokeos. [CORGA: CC2006-06/44]

Nunca me deu por pensar nese punto, quizabes porque foi tanto logro para min quitalo de diante, que vin aquela vitoria como única e rotunda, pero algo me falta por saber agora que me paro a pensar. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Agora ben, que ocorre cando o valor de *agora que* non é este senón o adversativo? É cando menos intrigante que a variante *ora que*, que non aparece no valor temporal, si o faga tamén para introducir, en palabras de Freixeiro (2005b: 211), “*o argumento máis forte nunha estrutura en que se opoñen dous xuízos*”.

No corpus de adestramento non apareceu aínda ningún caso destes, mais si os hai no CORGA, a pesar de que non son moi habituais. Maioritariamente encabezan secuencia, pero tamén se atopa esporadicamente algunha ocorrencia en posición interior. Vexamos algúns exemplos de cada.

A) Inicio de enunciado:

Agora que iso alá vostede. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Agora que, pregunte, ó mellor... [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Agora que, se ma deixa aquí, agradeceríallo. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Agora que mellor sería terlle preguntado ó Xeneral, para non perder o tempo como o está perdendo nestas lérias. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Agora que, fai o que vexas. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Agora que alá ti. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Un non vén cando quere senón cando pode, e se cadrou coa morte do vello, pois cadrou. Agora que eu non penso ir velo. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Que ben me soltou o conainas de Miro se viña para o enterro; que vaia sorna ten o cabrón; agora que eu tamén lle souben contestar, así coma quen é medio parvo, que quen era o morto, que como dicía o vello, non é máis listo quen parece, que a listura non se demostra, aparece. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Ora que, verbo destas entrevistas, cómpre sinalar que na maioría se entrecruzan coas que foi preciso inutilizar por non aproveitables. [CORGA: M. Gondar Portasany, *Mulleres de mortos. Cara a unha antropoloxía da muller galega*, Xerais, 1991]

Ora que tampouco fixemos reunións formais nin ningunha discusión a fondo sobre o tema. [CORGA: S. Álvarez, *Memoria da guerrilla*, Xerais, 1991]

Ora que como eramos moitos e a Policía seguía a vixia-la rexión, decidimos pasar unha tempada fóra. [CORGA: S. Álvarez, *Memoria da guerrilla*, Xerais, 1991]

E cando hai mobilizacións sempre sae el polo medio. Ora que cando xa está todo mobilizado. [CORGA: F. Alonso, *Trailer*, Xerais, 1992]

Ora que non me enteira eu de que ese vincheiro che pon a man encima... [CORGA: B. Iglesias Araúxo, *Vento de seda*, Xerais, 1992]

Nunca denantes se sentira como agora se sentía; ora que non sabería explicar, por moito que quixese, como era esa sensación. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

B) Interior de enunciado²⁰³

Tardaron máis dun ano en atopalo e o que trouxeron ao cabo, dicindo que era Stevo, foi un salvaxe do deserto, unha inmundicia de pel escura e pelo branco; para min que aquilo non podía ser, agora que se as análises das secuencias xenéticas o din, e segundo vostedes, non poden equivocarse, pois sería... [CORGA: R. Caride Ogando, *Soños eléctricos*, Xerais, 1992]

Foi el a man do destino, aínda que nada soubera, el foi o vencello do odio, agora que ben mirado, se non mo trae el atoparían outro que o trouxera, igual pensou que me facía un ben, si, é mellor pensar niso e esquecelo, ó cabo este sistema ensinoumo el. [CORGA: F. X. Fernández Naval, *Tempo de Crepúsculo*, Xerais, 1993]

Farei un xantar estupendo, un magnífico xantar, a base de marisco de luxo e froitas tropicais, verás como quedas contento, parece mentira, ¿non?, tantos anos convivindo e nunca tivera o detalle de convidalo a xantar un menú cociñado por min, sempre de bar en bar, ou de casa de amigo en casa de amigo, pratos quentes e pratos fríos, para a reuma non hai nada mellor que comer quente e non andar á chuvia, ou mesmo algo de licor, un licor non moi cargado, o alcohol xa se sabe, agora que, se non é reuma o que tes, ao mellor é unha libidinose, ollo con ela, disque os desequilibrios sexuais poden provocar desarraxos no hipotálamo capaces de dexenerar en patoloxías do máis extravagante con tal de non transparentaren o verdadeiro problema, curioso, ¿non?, eu sei algo diso, de cando cociñaba no Hospital Xeral, que alí todo eran conversas de moita altura, cos médicos, que sempre viñan petiscar algo entre horas e falaban do que lles chamaba a atención no seu oficio, unha verdadeira escola da vida, o Hospital, ¿sabes? [CORGA: M. Lourenzo González, *Noites de papel*, Xerais, 1997]

Rosiña: Eu non a escoito, que se cadra é o demo, agora que se oio algo... [CORGA: S. de Toro, *Unha rosa é unha rosa (Unha comedia de medo)*, Xerais, 1997]

²⁰³ Son os únicos casos que documentamos de *agora que* nos case 26 millóns de formas da versión 1.5 do CORGA, fronte aos 13 de *ora que*.

Segue, pois, habendo niveis de enerxía, os mesmos que obtivera Bohr, ora que non existe ningún modelo mecánico do átomo. [CORGA: A. Castelo Santaballa, *Temas de Química (COU da Universidade de Galicia)*, Edicións do Castro, 1989]

o que se consegue con isto é poder desimplicarse da situación ora que sen se ver obrigado a enfrontarse co instinto da vida. [CORGA: M. Gondar Portasany, *Mulleres de mortos. Cara a unha antropoloxía da muller galega*, Xerais, 1991]

Non nos consta que ó remata-la primeira guerra mundial reanudase as súas actividades esta Asociación, ora que isto non era atranco para que persistise o interese por Vigo como centro turístico. [CORGA: G. González Martín, *Vigo. No espello dos nosos avós*, Xerais, 1991]

a totalidade do proceso político non pasa polos partidos políticos, ora que si unha parte moi importante; [CORGA: M. Fraga Iribarne, *Da acción ó pensamento*, Ir Indo, 1993]

"_Escoitei as vosas frutíferas andanzas, coerentes algunhas pero as máis delas extraordinarias e pouco críbeis, ora que gustosamente todas distraídas. [CORGA: L. C. Carballal, *Historias non irrealis*, Espiral Maior, 2001]

Este puido co vinilo, si, ora que non totalmente. [CORGA: CC2006-01/32]

Si, é certo, as cartas escritas a man non son tecnoloxía, ora que constitúen se cadra, a parte máis feble da vida tradicional fronte ós avances da era dos díxitos. [CORGA: CC2006-01/35]

Evidentemente, o significado denotado por *agora* na combinación con *que* que vimos de ver non é o de valor temporal deíctico de momento presente ou inmediatamente antes ou despois dese momento, senón o adversativo. Con todo, o adverbio *agora* non é alleo a este outro significado, e tanto é así, que o reproduce tamén na locución *agora ben, ora ben* e por si só ou na variante *ora*:

Esas clases que se dan na universidade, na que o propio complexo de inferioridade do profesor trata de acumular datos, citas, bibliografía, un pouco como para parapetarse, iso alí non existía; citaban o imprescindible, agora, iso si, demorábanse no encerado o tempo que fixese falla e volvíase unha e outra vez sobre o explicado, se alguén non o entendía. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

Non me arrepinto. Dende logo, é o que ten máis datos reais e biográficos de tódolos que están no libro; agora, arrepentirme, de ningún. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

En España empezáronse a ver películas españolas na televisión con Méndez Leite; antes, España enteira facíalle o boicot e os cines puñanas porque lles obrigaba o Goberno; agora, en canto puñan unha norteamericana, enchíanse os locais. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

"Eu pagarvos non podo; agora, se queredes tocar, por iso non hai problema." [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]

"Agora ben, a postura oficial _continuou_ vai ser apostar pola estabilidade no emprego e non abarata-lo despedimento". [CORGAetq: CG1999-05-04/55]

Agora ben, é sincero todo este sentimento humanitario a prol do pobo de Darfur? [CORGAetq: GH2008-08-05/97]

Desta forma o home vive nunha organización máis ou menos racional do mundo social, agora ben, vive inxenuamente. [CORGA: X. Cabrera Varela, *O combate dos deuses. Unha introducción a Max Weber*, Xerais, 1990]

Trátase de ter máis coches, TV, camións, autoestradas, electricidade, teléfonos, comercios..., agora ben, cabería preguntarse se todo o producido mellora o nivel de goce da comunidade contribuíndo ao seu benestar, ou ten outras consecuencias. [CORGA: L. Arias Rodríguez e R. Prieto-Puga Fariña, *Economía. Primeiro curso de Bacharelato*, Bahía, 1998]

Resulta necesario dicir, mal que lle pese a alguén, que cada un destes tres sistemas de distribución teatral representa un estadio máis avanzado que o anterior e corresponde a unha fase de normalización teatral superior, agora ben, nada apunta a considerar que sexan incompatibles entre si, é máis, posiblemente un sistema que combinase os tres resultaría dunha grande eficacia: [CORGA: TDN1997-01/1]

Eu non quería que naide sospeitase de min neste senso. Ora, si me vían entrar todas as mañás na Comisaría, aparentemente pola miña propia vontade, era fÁCile supoñer que eu podía ser un dos denunciante. [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Mentres dura a reacción o sistema $Zn+H_2SO_4$ será, pois, aberto. Ora, se poñemos un tapón que non deixe fuxi-lo hidróxeno, o sistema será pechado. [CORGA: A. Castelo Santaballa, *Temas de Química (COU da Universidade de Galicia)*, Edicións do Castro, 1989]

Pois si, na Europa Central (e do Leste tamén!) son afeccionados a esta serie, e supoñemos que iso non é nin bo nin malo, ora, o que si que é especialmente grave son: [CORGA: CC2006-01/39]

O xenocidio é a realidade innegábel coma acontecera nos días derradeiros dos Pahalevi en Irán. Ora, ¿por que xornais editados aquí deron por boa durante tanto tempo a decisión das grandes axencias de chamar baixas aos mortos da Intifada e asasinados aos soldados de Sharon? [CORGA: ANT2002-09-26/43]

Quedei petrificado ao volver ver de golpe o campo cheo de persoas, xente e máis xente a saltar o valado e penetrar no recinto, quizais porque o bruar volvía exercer a súa atracción poderosa sobre todo aquel que o escoitaba. Ora, desta volta, os que ían penetrando no recinto parecían decididos a lle poñer remate a aquela estraña condena coa determinación de botar ao chan cada unha das pedras daquelas construcións. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Ora ben, a cuestión non é tan simple. [CORGA: R. Villares, *Foros, frades e fidalgos*, Xerais, 1982]

Ora ben, o tráxico suceso da chamada "casa dos cataláns" (as instalacións da Cía Gasset e Botana) non foi un acto primitivo e animal de vinganza, anque primitivo e animal sexa. [CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

Ora ben, as cartillas agrícolas non constituían o medio máis axeitado para acadar este obxectivo. [CORGA: N. de Gabriel, *Agricultura e escola. Contra a rutina e o éxodo rural*, Universidade de Santiago de Compostela, 1989]

Ora ben, o xurelón sabe de tódalas maneiras: frescal, asado, nas brasas ou cocido, con aceite en cru e cebola. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

Ante as dúbidas que nos suscitaba a categoría gramatical das unidades *agora*, *agora ben* e *agora que*, xunto, por suposto, coas súas variantes *ora*, *ora ben* e *ora que*, decidimos facer unha peregrinaxe polas distintas obras lexicográficas galegas máis recentes e confrontar a información obtida con algún dicionario do portugués e tamén do castelán. A conclusión que se pode extraer de tal consulta, ante a falta de homoxeneidade e as contradicións existentes, como veremos deseguido, é a lexitimación da dúbida categorial.

Resumimos a continuación a información obtida para as unidades *agora/ora* e as expresións *agora/ora ben* e *agora/ora que*, mais, iso si, prescindimos de recoller ningunha acepción para *agora* que non reflecta o significado adversativo ou disxuntivo-distributivo.

A) GALEGO

DRAG, 1997:

Agora ben. *loc. conx.* Introduce unha oposición ou puntualización a algo que se dixo anteriormente. *Eu aconsélloche que vaias; agora ben, fai como queiras.* SIN. **emporiso**, **ora ben**.

Agora que. *loc. conx.* *Pero. É un rapaz intelixente, agora que algo preguiceiro.* SIN. **ora que.**

Ora. *conx.* Indica alternancia ou distribución; unhas veces... outras... OBS. Úsase repetida. *Ora el, ora seu pai, sempre estaba alguén comigo. Ora ría, ora choraba.* SIN. **ben, quer, xa.**

Ora ben. *loc. conx.* Aceptando isto, non impide que [cousa que se indica]. *Fun eu, ora ben, non o fixen a propósito.* SIN. **agora ben, emporiso.**

Ora que. *loc. conx.* Agora que, pero. *Acabou por ir, ora que non quería.*

DRAG, 2012²⁰⁴.

Agora ben *loc. conx.* Introduce unha oposición ou puntualización a algo que se dixo anteriormente. *Eu aconsélloche que vaias; agora ben, fai como queiras.* SIN. **emporiso, ora ben.**

Agora que *loc. conx.* Introduce unha restrición ou unha oposición a algo que xa se dixo. *É un rapaz intelixente, agora que algo preguiceiro.* SIN. **mais, ora que, pero¹.**

Ora *conx.* Indica alternancia ou distribución; unhas veces... outras... OBS. Úsase repetida. *Ora el, ora seu pai, sempre estaba alguén comigo. Ora ría, ora choraba.* SIN. **ben, quer, xa1.**

Ora ben *loc. conx.* Aceptando isto, suposto isto. *Fun eu, ora ben, non o fixen a propósito.* SIN. **agora ben, emporiso.**

Ora que *loc. conx.* Agora que. *Acabou por ir, ora que non quería*

Diccionario de IR INDO, 2004:

Agora ben *loc conx* Introduce unha pequena oposición ou puntualización a algo xa dito.

Agora que *loc conx* Indica a oposición entre dous elementos ou membros que une.

Ora *conx* Indica dúas posibilidades, anteposto a dous elementos.

Ora ben *loc conx* Introduce unha pequena oposición ou puntualización a algo xa dito.

Ora que *loc conx* Marca a oposición entre dous elementos ou membros que une, de xeito que o que se di no segundo é unha restricción do primeiro.

Diccionario SÉCULO21, 2006:

Agora 5. *conx ad.* Pero, non obstante

Agora ben *loc. adv.* Emporiso, tendo en conta

²⁰⁴ O noso agradecemento ao Seminario de Lexicografía da Real Academia Galega e, en especial, a Ana e Pazos, pola rapidez, amabilidade, paciencia e dispoñibilidade que sempre demostraron ante as nosas consultas sobre que información ía proporcionar a nova edición do diccionario da Academia. [Consulta realizada o 29/02/2012]. Na consulta á versión en liña, realizada o 30/06/2013, verificamos que non se produciu ningún cambio na definición destas voces.

Agora que. *loc. conx.* Pero, con todo.

Ora. *conx. dis, tr.* Anteponse a dúas palabras, frases ou oracións que expresan dúas cousas ou posibilidades (*ora si, ora non*). Sin. *ben, quer, xa*. **Ora ben** Agora ben. **Ora que** Agora que.

Diccionario XERAIS, 2009:

Agora ben *loc. adv.* Denota que, unha vez admitido o anterior, hai que considerar o que segue; con todo, así e todo.

Agora que. *loc. conx.* Pero, aínda que.

Ora *conx.* 1. En frases e oracións distributivas, e acompañando as dúas proposicións, expresa dúas posibilidades (*acudiremos a vela, ora hoxe, ora mañá*).

Ora ben *loc. conx.*, pero.

Ora que *loc. conx.*, pero tamén.

Só Século21 se desmarca da tendencia xeral ao recoñecer unha acepción para *agora* que non é a de adverbio temporal, senón a adversativa, que cataloga como conxunción. Tamén Século21 e Xerais distinguen categorialmente as dúas locucións: *agora ben* adverbial, *agora que* conxuntiva, malia que o último é incoherente á hora de catalogar a variante con *ben* completa (*agora ben* = adverbial) e a reducida (*ora ben* = conxuntiva).

B) CASTELÁN

Para o castelán obtivemos tamén datos contraditorios:

Diccionario da RAE, edición en liña²⁰⁵:

Ahora 6. *conj. advers.* Pero, sin embargo

Ahora bien *loc. conjunt.* Esto supuesto o sentado. *Ahora bien, ¿qué se pretende lograr con esa diligencia?*

Agora que *loc. conjunt.* pero (para contraponer un concepto a otro). *La casa es cómoda, ahora que no tiene ascensor.*

Ora (Afér.) *conj. distrib.* ahora. *Tomando ora la espada, ora la pluma.*

Diccionario MARÍA MOLINER, 2007:

Ahora... ahora... Expresión disyuntivo-distributiva: *Ahora llore, ahora ría...*

²⁰⁵Dispoñible en <http://www.rae.es> [consulta realizada o 29/02/2012].

Ahora bien Equivale a «pero» en expresiones como "Haz lo que quieras; ahora bien, atente a las consecuencias".

Ahora que 1. Pero. "Es verdad lo que dices; ahora que eso no es todo". 2 Se emplea con verbos como *ocurrirse* o *acordarse* en frases que inician la exposición del pensamiento que se ocurre o se recuerda. "Ahora que me acuerdo, me dio un recado para ti" "Ahora que pienso, será mejor que dejemos la visita para mañana".

Ora conj. Afér. de «ahora», usada solamente como conjunción distributiva, repetida, en frases literarias de sabor anticuado: *Tomando ora la espada, ora la pluma*.

Diccionario SECO, 1999²⁰⁶:

Ahora bien. Denota que, una vez admitido o sentado lo anterior, hay que considerar la puntualización que sigue. Tb, simplemente, *ahora*; en este caso, la oración que sigue puede ir iniciada por QUE. ... *Trataba bien a la gente que tenía. Ahora, eso si, con distancia, como era él. Los entendidos ... dicen que para febrero de 1975 será el momento de máxima depresión. Ahora que, siempre según los expertos, a partir de febrero la coyuntura aflojará la tendencia negativa.*

Ahora... ahora... (lit) Unas veces... otras.... *El humor del general se envileció: se manifestaba ahora agrio y silencioso, ahora en exceso alegre y jovial.*

Ora. ~..., ~... (lit) **I adv** 1 Unas veces..., otras veces.... *¡El afilador! En sus piernas se teje, paso a paso, el tapiz de la patria, ora pardo y yermo cual páramo .., ora húmedo e industrial cual la verde Vasconia, ora lumínico y soleado cual la Andalucía.* **II conj** 2 Tanto si... como si... Con v en subj. *Quedaban .. detenidos en el gesto y en el tiempo para dar gloria a Dios, "ora bebáis, ora hagáis cualquier cosa".*

C) PORTUGUÉS

Diccionario Novo Aurélio, 1999:

Agora Conj. 7. Mas, porém, contudo, todavia: *Ir é fácil; agora, voltar é que são elas.*

Ora2. Conj. 1. Mas; note-se (que): *De repente chegou ele: ora naquele dia eu estava impaciente, e não o atendi como era devido;* "Ora, se deu que chegou / (isso já faz muito tempo) / no bangüê dum meu avô / uma negra bonitinha / chamada negra Fulô." **Adv.** 2. Agora; atualmente, presentemente: *Meu amigo Paulo, que ora se acha aqui, deseja falarlhe.*

Diccionario do português de Brasil, 2002:

Agora Conj [Coordenativa. Adversativa] 11. introduz um elemento que contraria uma expectativa; mas: *Arrume-se o melhor que puder e quando ele chegar eu faço as*

²⁰⁶ Mantén ambas as expresións baixo a categorización de *ahora* como adverbio.

apresentações. Agora, o resto é com vocês. **12.** introduz um elemento que se acrescenta ao que foi dito, mas con expectativa de resposta contrária: *Por que você não atira dinheiro no auditório como o Sílvio?/ -Porque eu não tenho / -Agora, você acha eticamente incorreto, acha que é ofensivo jogar grana pro público?*

Ora Adv. [Tempo] 1. neste momento; agora (...) . [Inclusão] 2 expressa uma transição de pensamento, en continuación ao discurso anterior; demais; além disso: *Que estariam discutindo eles, absortos a ponto de não atentarem para os lados? Tia Gilda. Ora, não se podia dizer que tia Gilda fosse, a rigor, uma bela mulher.* Conj. [**Coordenativa. Alternativa**] 5 ou ou; já já; quer quer: *Parecia, naquele rouquejar arfante, lançar queixas contra os brancos, ora no dialeto camaiurá, ora num português errado.* [**Adversativa**] 6 mas: *Tou derrubado, homem, e hoje não podemos fazer negócio/ -Ora, isto é maleita, Catarino, e eu te curo à toa.* 7 mas: *Por enquanto é uma idéia. Não sei nada ainda de partido. Não quero voltar ao meu também. Ora, você sabe que os partidos políticos estão atrás dos cargos, mais nada.*

Deixando de lado a especificación categorial das locucións, sorprende nos dicionarios galegos a igualdade descrita para os pares *agora ben ~ ora ben* e *agora que ~ ora que*, fronte ás diferenzas que amosan para as correspondentes formas simples: *agora* e *ora*, para as que, coa salvidade xa mencionada de Século 21, *agora* é un adverbio e *ora*, unha conxunción distributiva, ignorándose por completo deste xeito o valor adversativo que ambas as formas soas poden amosar e mais o posible valor distributivo de *agora* e o adverbial de *ora*.

Segundo Freixeiro (2005b: 211-217) os conectores que expresan a noción de preferencia (*antes (ben)*, *ora (ben)* e *máis ben*) proxectan a idea de preferencia, superioridade ou prioridade metaforicamente no plano textual para denotar a adversatividade, ou sexa, superioridade dun argumento sobre outro. É interesante o razoamento que sobre a súa gramaticalización nos ofrece este autor:

Estes conectores teñen en común o feito de procederen de termos que implican a presenza doutro termo, por regra xeral anterior, con que se achán dalgún modo relacionados: *antes* apunta a algo previamente mencionado, (*ag*)*ora* adoita estar conectado a un 'antes' e *máis ben* subliña a superioridade dun termo en relación a outro anteriormente presentado; así, antes da súa gramaticalización como partículas adversativas, estabelecían unha relación entre dous fragmentos do texto. (...) Alén diso, como nas estruturas en que se orixinaron existía unha noción de contraste (entre un antes e un despois, entre algo mellor e algo peor), isto tivo de favorecer que *antes ben*, *ora ben* e *máis ben* acabasen por adquirir un significado contraargumentativo. [Freixeiro, 2005b: 212]

Así como tamén resulta moi interesante a descrición específica para *ora(ben)*:

Deste valor continuativo de *ora ben* talvez derive o seu sentido adversativo: o conector sinala a validez do enunciado anterior e a maior relevancia do que vén a seguir, marcando por tanto a preferencia polo segundo argumento. Como a noción de preferencia dun argumento é básica nas partículas adversativas, será suficiente con que apareza en contextos en que se exprese contraste para que *ora (ben)* se converta en conector adversativo.

Porén, existe unha clara diferenza con *antes (ben)* e *máis ben*: *ora (ben)* ten un significado adversativo restritivo, non exclusivo, derivado do seu valor continuativo, que non supón exclusión de ningún xuízo: isto é, expresa a preferencia por un argumento sen que iso supoña invalidar o menos pertinente, de modo que a relación entre os dous non é de incompatibilidade. (...) Fronte a outros conectores contraargumentativos parentéticos, *ora ben* carece de mobilidade, debendo ocupar sempre a posición inicial do segmento que introduce, como acontece con *antes ben*. Por outra parte, *ora ben* é un conector que habitualmente é monolgal, pois une os enunciados dun mesmo locutor, ao contrario de *mais (e pero)*, que pode aparecer nos dous contornos, o monolgal e o dialogal; igualmente pode unir non só informacións oracionais, senón tamén parágrafos enteiros, semellando ser esta a súa función habitual. [Freixeiro, 2005b: 215-216]

O que máis nos sorprendente na cita que acabamos de recoller deste filólogo e gramático, precisamente por ser ese o noso interese principal, é que non clasifica categorialmente *(ag)ora* nin *(ag)ora (ben)*. O autor emprega o termo xenérico *partícula* ou o específico *conector discursivo*, mais non explica se o conector se adscribe na mesma clase na que supostamente ten a súa orixe, o adverbio *agora*, nin aclara a formación con *ben*, nin analiza as diferenzas, se é que as hai, entre a forma simple e a composta.

Si se pronuncia Freixeiro no seu estudo parcial sobre os conectores contraargumentativos (2005a: 252), pero faino só sobre a categoría de *ora ben* para quen remite á adverbial.

Nós imos partir da análise sobre a posible clasificación das formas simples para chegar logo a unha posible delimitación categorial das locucións nas que estas están inmersas.

Non podemos perder de vista que unha, *ora*, é a aférese da outra, *agora*, polo que cobra sentido que os valores dunha poidan ser desempeñados pola outra ou que se acabe producindo unha especialización entre elas.

Comezaremos polo uso distributivo-disxuntivo de *ora* e tentaremos ver se este se lle pode aplicar tamén a *agora*. Fixémonos primeiro en *ora*. Aparece amosando distribución con distintas unidades: tanto con cláusulas, como con frases nominais, preposicionais e adverbiais. Por exemplo:

Todo o teu corpo se move, inclínase, ora á dereita, ora á esquerda; [CORGA: F. Alonso, *Trailer*, Xerais, 1992]

Martiño: Ora queimaba o sol, ora chovía, ora abafaba a calor, ora roía o frío, ora ventaba, ora... [CORGA: C. Pazó, *Raiñas de pedra*, Xerais, 1994]

Leva ás costas un saco de ambicións que ora o sepultan baixo o seu peso, ora lle serven de escada inesperada. [CORGA: *TN1997-06/13*]

Directamente relacionadas coa implantación das centrais de reutilización e incineración de Cerceda están as famosas empacadoras que, ora aquí ora aló, encirran os ánimos dos veciños de lugares que non queren espertar cheirando a lixo. [CORGA: *TN1997-06/17*]

Estes inventos dos que se valeron foron a artillería e a imprenta, que se aunan dalgún xeito nese ideal do home renacentista que ben toma "ora a pluma, ora a espada". [CORGA: *P1997-04/4*]

Xiz estaba sumido nun estado de resignación paralítica, a penas os ollos eran quen de se mover de órbita, ora aquí, ora na distancia; [CORGA: X. Vázquez Pintor, *Quen faga voar*, Laivento, 2000]

Non son moitas as ocorrencias no CORGA, mais algunhas si hai, nas que *agora* actúa como elemento conxuntivo que marca unha estrutura discontinua distributiva. Vexamos algúns exemplos:

Verdadeiramente resúltache un home ben extraño este Miguel de la Torre, que te agarra agora amistoso polo brazo e quere arranxa-la situación agora con palabras novas, porque xa abonda de bromas e hai que chegar ós feitos coma amigos. [CORGA: Ú. Heinze, *O soño perdido de Elvira M.*, Xerais, 1982]

Era un deses firmamentos tormentosos que tanta impresión causan e que ameazaba con ser totalmente distinto de minuto en minuto. Agora lucía o sol por entre unhas nubes galopantes, agora unha sombra inmensa sumía a rúa enteira nunha especie de preludio nocturno. [CORGA: M. Forcadela, *Barato, barato*, Sotelo Blanco, 1993]

Agora está aquí, agora está alá... [CORGA: M. Seixas, *A velocidade do frío*, Xerais, 1996]

Adiviñábase, máis non se vía, que as folepas caían mestas, seguidas, abondosas, e, ollando fixo o seu vagar, víanse subir e baixar empuxadas polo vento, agora como danzando, agora

súpeto rápidas, grandes, decididas, igual que caerían se nunca máis fose deixar de nevar. [CORGA: X. Miranda, *Infancia e desventura de Lino Carrán*, Galaxia, 2000]

Ó outro lado dos pés de micro, no mar escuro e fondo do público, os lampexos da luz sobre a brétema branca deixan ver por momentos, agora aquí, agora acolá, as caras da xente que esta noite veu compartir isto con nós. [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]

Home 2: ¿E se pensan que cambiamos de chaqueta? Agora á dereita, agora á esquerda. Será suficiente para que disparen, tanto se son roxos como nacionais. [CORGA: G. Pernas Cora, *Medidas preventivas*, Biblos clube de lectores, 2006]

Xenebra: Como aves, nós tamén migramos, agora maitines, agora tercias, agora completas. [CORGA: T. Moure, *Unha primavera para Aldara*, Deputación Provincial da Coruña, 2008]

É verdade que en todos eles *agora* mantén o carácter de adverbio de tempo, mais tamén é certo que os membros poden ser máis de dous (último dos exemplos) e que ese mesmo carácter de adverbio temporal tamén o amosa *ora*, para a que, nestes casos, non se discute a súa categoría como conxunción distributiva-disxuntiva. Nós cremos que nos enunciados anteriores *agora* non mostra carácter deíctico (oposto a *daquela*, *naquel momento*) e que prima neles a expresión distributiva, polo que categorialmente consideramos que estamos ante unha conxunción discontinua, a semellanza de *ora... ora...*

O seguinte paso na nosa análise é comprobar se *ora* aparece tamén como adverbio temporal. Para iso rastrexamos o CORGA, pois non nos interesa tanto confirmar que o foi nalgún momento da historia da nosa lingua como verificar se no galego contemporáneo se usou/usa como variante de *agora*. Naturalmente, debemos buscar secuencias nas que se poida contrastar, implícita ou explicitamente, cun adverbio temporal pasado (*antes*, *entón*) ou futuro (*despois*) para que non quede lugar a dúbidas. O que atopamos non é moito. Vexamos unha mostra:

O Vello: O mesmo, señor, o mesmo, así é o meu coello. Se por algures o veu xa sabe que ese é meu. [O Vagabundo:] Se foi seu xa non o é, ora non é de ninguén, / Ver, vino e papeino con arroz e con cenorias, un cacho comino asado, o resto en pepitoria. [CORGA: A. Cortizas Amado, *O coleccionista de sombras (Farsa en tres actos con dous entreactos)*, Sotelo Blanco, 1990]

E de tal arredor emerxeron como anémonas os corcosidos, ora con harpa e despois con laúde, pequenos reis de corredor e milagres, á procura do maseirín e outras sinecuras, erguendo no ar

putrefacto "a rebelión das masas", "a nación invertebrada", o bourco, o rebumbio. [CORGA: CG2001-10-02/8]

Os casos atopados son poucos, o que lles confire case un valor anecdótico, pero veñen demostrar as nosas sospeitas: se un é aférese do outro, tiveron/teñen que comparecer nalgún momento da historia nos mesmos contextos e coas mesmas funcións. O corpus demostra que aínda o fixeron recentemente, inda que polas frecuencias non en igualdade de condicións, xa que parece que a lingua buscou a especialización: *agora*, adverbio e *ora*, conxunción. Naturalmente, se algún dos exemplos anteriores xurdise nun texto a desambiguar manualmente, a etiqueta que lle asignariamos sería a de Adverbio.

O seguinte chanzo que debemos subir, cremos, é tentar dilucidar que categoría lles corresponde aos *ora*, *agora* adversativos. Estes, igual que sucede coas locucións *ag(ora) ben* e *ag(ora) que*, presentan información novidosa que supón orientación argumentativa contraria á que os datos precedentes suxiren. Recollemos só un par de exemplos de cada dos recollidos máis arriba nesta mesma epígrafe para telos próximos:

Quedei petrificado ao volver ver de golpe o campo cheo de persoas, xente e máis xente a saltar o valado e penetrar no recinto, quizais porque o bruar volvía exercer a súa atracción poderosa sobre todo aquel que o escoitaba. Ora, desta volta, os que ían penetrando no recinto parecían decididos a lle poñer remate a aquela estraña condena coa determinación de botar ao chan cada unha das pedras daquelas construcións. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Pois si, na Europa Central (e do Leste tamén!) son afeccionados a esta serie, e supoñemos que iso non é nin bo nin malo, ora, o que si que é especialmente grave son: [CORGA: CC2006-01/39]

En España empezáronse a ver películas españolas na televisión con Méndez Leite; antes, España enteira facíalle o boicot e os cines púñanas porque lles obrigaba o Goberno; agora, en canto puñan unha norteamericana, enchíanse os locais. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

"Eu pagarvos non podo; agora, se queredes tocar, por iso non hai problema." [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]

A nosa dúbida redúcese á categoría gramatical na que debemos adscribilos. Adaptando os criterios de Zorraquino e Portolés (1999) e Portolés (2001) ao galego, o que tamén realiza Freixeiro (2005a), imos relacionar deseguido unha serie de

coincidencias e diferenzas entre as dúas categorías que nos provocan a dúbida (Adverbio fronte a Conxunción) e veremos se damos chegado a unha conclusión.

Así, partindo do feito de que *agora e ora* son unidades invariables, o que é característica común a adverbio e conxunción, comprobamos que:

1. Na acepción contraargumentativa non admiten negación nin modificadores; non obstante, por unha parte, esa tamén é unha peculiaridade de todas as conxuncións e, pola outra, dos adverbios que son marcadores dicursivos.

2. Dúas conxuncións non vinculan sintacticamente as mesmas unidades, de modo que non se documentarán as combinacións *e pero* ou *aínda que pero*, mais si é habitual a coincidencia dunha mesma conxunción e dun adverbio conector nun mesmo enunciado.

O galego, igual ca o castelán, admite a combinación redundante pero enfática dunha conxunción adversativa –xeralmente *pero* ou *mais*– e varias locucións adverbiais tamén de sentido adversativo: *sen embargo, non obstante, en cambio, antes ben, etc.*, mais isto non sucede, ou polo menos non o documentamos, con *agora/ora* nin *agora ben/ora ben*.

Era verán, corría o mes da sega, pero, non obstante, a mesa estaba invadida de mariscos: percebes, camaróns, ameixas, ostras, nécoras, centolas, lumbrigantes, cigalas... [CORGA: X. Costa Clavell, *O conde de Viloíde*, Sotelo Blanco, 1995]

Respecto dos soportes que suxeitaban os anuncios máis denunciados, a televisión volve encabeza-la listaxe (32,9%) pero destacan, non obstante, este ano o medio exterior (valados, mobiliario urbano e autobuses) que en conxunto representan o 30% cando o ano pasado non chegou ó 9%. [CORGA: CG2002-05-31/41]

Os elementos Si e Ge posúen a mesma estrutura cristalina que o diamante pero sen embargo son semicondutores, o que non é de estrañar pois o carácter metálico aumenta ó descender en calquera grupo da Táboa Periódica. [CORGA: A. Castelo Santaballa, *Temas de Química (COU da Universidade de Galicia)*, Edicións do Castro, 1989]

Estaba claro matinou o que tiña que facer, pero en cambio non o estaba o que debía facer en tanto o que tiña que facer non se podía facer. [CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

Para Weber o continuum heteroxéneo da realidade non consiste nun fluxo de sensacións, senón, antes ben, nunha realidade fenoménica xa constituída. [CORGA: X. Cabrera Varela, *O combate dos deuses. Unha introducción a Max Weber*, Xerais, 1990]

como no Terceiro Mundo, o obxectivo deses programas non era alenta-lo ben-estar económico senón, antes ben, garanti-la lucrativa explotación da periferia do sistema capitalista mundial. [CORGA: C. Taibo, *Europa sen folgos. Para entendérmo-los problemas das periferias europeas*, Sotelo Blanco, 1994]

Non obstante, si documentamos tres ocorrencias nas que o *agora* adversativo vai seguido das locucións adverbiais contraargumentativas *con todo* ou *iso si*:

Cada vez que hai probas de motociclismo no Jarama, o trailer de Pepe é adiantado por enormes greas de motos que pasan pola estrada asubiando coma mísiles Tomahawk ou cazabombardeiros F-15. O asfalto convértese daquela nun vómito interminable de motos de arrepiante potencia. Agora, con todo, as inclemencias meteorolóxicas vencen ás máquinas. [CORGA: F. Alonso, *Trailer*, Xerais, 1992]

(Con Carrascal pode que non nos decatemos moito das noticias, agora, iso si, estamos facendo unha cultura de idiomas tan guapamente, e ademais gratis). [CORGA: *CG1996-02-25/141*]

Si, moito vacilar co meu con Raquel, pero... ¿e el que? Porque moito fardar disto e mais do outro, pero que eu soubese... mellor dito, que eu tivese algunha evidencia... Agora, iso si, para quen lle prestase oídos un par de veces, o meu camarada tiña de ser, tal como se expresaba e conducía, todo un tipo experimentado nos máis complexos saberes; [CORGA: X. R. Pena, *A era de Acuario*, Xerais, 1997]

E malia que é moi infrecuente, tampouco sería o único caso no que aparecen dúas locucións adverbiais de matiz adversativo:

Sen embargo, con todo... Tiña que facelo. Aspirou con ansia o fume, deixou a mente en branco á busca dunha saída. [CORGA, X. Alcalá, *O Larvisión e outros relatos*, Xerais, 1984]

3. Así mesmo, as conxuncións sitúanse obrigatoriamente entre os dous membros que unen. mentres que os adverbios conectores, polo contrario, adoitan posuír unha maior mobilidade e apareceren en forma parentética.

O problema xórdenos aquí porque, segundo Portolés (2001) e nós confirmámolo, resulta moi difícil documentar algúns marcadores adverbiais nunha posición que non sexa a inicial do seu membro discursivo, só precedidos, no seu caso, por unha

conxunción. Isto sucede con *agora ben, a propósito, a saber, pois ben, antes ben* e algún outro²⁰⁷.

De todas as ocorrencias de *agora, ora, agora ben* ou *ora ben* que atopamos, só documentamos dúas aparicións en posición medial, parentética, separando o Predicado do Complemento Directo, é dicir, comportándose coma unha locución adverbial discursiva calquera. A nosa percepción como falantes dinos que son casos anómalos, mais aí están:

Iso non lle impedía, ora ben, estar a ben con el, o que sempre era conveniente traballando no Berbés. [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

Esta enerxía nuclear, contrariamente ao que se podería pensar pola reducida escala á que se atopa, resulta enorme en relación coas outras enerxías coñecidas, para unha mesma cantidade de material e, tamén, máis perigosa. Existen outras manifestacións da enerxía algo máis difíciles de representar seguindo o noso modelo mecánico e que veremos posteriormente: a enerxía eléctrica e a enerxía radiante. É doado observar, agora, que o café posúe unha enerxía térmica (por iso dicimos que está quente) maior que a do xelado, pero a enerxía química, almacenada nas estruturas moleculares que constitúen o xelado (principalmente graxas e azucres), pode ser moitísimo maior que a do café, maioritariamente formado por auga, así pois, será importante coñecer a súa composición química para dar unha resposta precisa. [CORGA: R. Vilalta, M. Cid e A. Varela, *Física e química*, A Nosa Terra, 1997]

Nas escasas restantes ocasións, a pesar de atopárense algúns exemplos entre comas, a puntuación existente nas secuencias non reflicte, na nosa opinión, a verdadeira entoación nin as pausas reais, xa que a inserción da partícula, que puntualiza e precisa a información anterior reorientando o discurso, require unha pausa marcada, tanto anterior coma posterior. Véxase unha pequena mostra:

Esas clases que se dan na universidade, na que o propio complexo de inferioridade do profesor trata de acumular datos, citas, bibliografía, un pouco como para parapetarse, iso alí non existía; citaban o imprescindible, agora, iso si, demorábanse no encerado o tempo que fixese falla e volvíase unha e outra vez sobre o explicado, se alguén non o entendía. [CORGA: X. M. Salgado e X. M. Casado, *X. L. Méndez Ferrín*, Sotelo Blanco, 1989]

²⁰⁷ Non é verdade que *antes ben* deba ocupar sempre a posición inicial como afirma Freixeiro (2005b), senón que se documenta tamén, aínda que esporadicamente, como matizan Zorraquino e Portolés (1999: 4064), en posición medial e separando funcións no grupo do Predicado:

A paz, de acordo con isto, debe ser considerada, antes ben, como unha conquista, non como un punto de partida e, menos aínda, como unha idade de ouro perdida. [CORGA: F. Sampedro, *A violencia excedente*, Laiovento, 2006] reivindica, antes ben, unha clara recuperación destes nun escenario marcado, iso si, polo rexeitamento do consumo irracional. [CORGA: *GH2009-04-26/125*]

Pois si, na Europa Central (e do Leste tamén!) son afeccionados a esta serie, e supoñemos que iso non é nin bo nin malo, ora, o que si que é especialmente grave son: [CORGA: CC2006-01/39]

_Home, haberaos ananos e de tódolos tamaños, supoño, eu que sei _dixo Ramiro_, ora, que os Mouros son unha cousa e os Ananos outra. [CORGA: X. Miranda, *Historia dun paraugas azul*, Xerais, 1991]

Trátase de ter máis coches, TV, camións, autoestradas, electricidade, teléfonos, comercios..., agora ben, cabería preguntarse se todo o producido mellora o nivel de goce da comunidade contribuíndo ao seu benestar, ou ten outras consecuencias. [CORGA: L. Arias Rodríguez e R. Prieto-Puga Fariña, *Economía. Primeiro curso de Bacharelato*, Bahía, 1998]

De 1994 a 1995 os casos aumentaron un 52%, ahora ben, unha chamada ó optimismo: os casos rexistrados entre xaneiro-marzo de 1996 foron un 59% menos que en igual período de 1995. [CORGA: J. J. Gestal Otero, *Enfermidades infecciosas emerxentes*, Xunta de Galicia, 1997]

Así pois, salvo o caso do *agora ben* e do *agora* mediais que se ubican entre o Predicado e o CD, ningunha outra das unidades que estamos estudando presenta a mobilidade característica dos conectores discursivos adverbiais.

4. Continuamos parafraseando a Portolés e Freixeiro: os adverbios conectores non posúen as propiedades sintácticas dunha conxunción coordinante, aínda que si poden relacionar pragmaticamente dous membros do discurso, de modo que son gramaticais construcións como *é feo, mais simpático* pero non **é feo, non obstante, simpático*, aínda que si *é feo, mais, non obstante, simpático*, pois a conxunción *mais* habilita como gramatical a relación entre dous adxectivos, autorizando de paso a aparición do adverbio.

Se as unidades en discusión fosen conxuncións, terían que resultar gramaticais as secuencias paralelas á do exemplo con *agora, agora ben, agora que* e as variantes con *ora*. Mais, o que obtemos é:

*É feo, agora simpático.

*É feo, agora ben simpático.

?É feo, agora que simpático.

Pola contra, se fosen adverbios, a anteposición da conxunción *pero* debería converter as secuencias en gramaticais:

*É feo, pero, agora, simpático.

?É feo, pero, agora ben, simpático.

*É feo, pero, agora que, simpático.

Nas secuencias con *agora*, en ambas as dúas, non asoma o sentido adversativo senón o valor temporal. Semella que a falta de corte co anterior, reflectido na ortografía con punto e coma, dous puntos e mesmo punto final ou á parte, é un requisito imprescindible para que o adverbio deixe de especificar temporalidade.

Do mesmo xeito, a falta da pausa tras *agora ben* seguido de adxectivo provoca que o *ben* se comporte como adverbio intensificador e forme grupo co adxectivo e non co *agora* que o precede.

Finalmente, a en teoría locución conxuntiva *agora que* non pode ir precedida, loxicamente, de conxunción *mais*, non obstante, tampouco debería poder estar separada dos dous membros que une por pausa.

Con todo, semella que as dúas secuencias marcadas cun interrogante non producen o rexeitamento que si provocan as sinaladas cun asterisco, polo que poderían ser gramaticais.

5. Ao contrario ca as conxuncións, os adverbios conectores pódense coordinar con outros sintagmas xeralmente situados entre pausas, coma en *Mais, non obstante –e a pesar de todo– conseguiu o emprego desexado*.

Nós non documentamos no CORGA ningún caso de coordinación entre algunha das unidades obxecto desta reflexión e outro sintagma do seu mesmo grupo significativo ou próximo. Como moito, o único exemplo que se aproxima un pouco ao que buscamos é unha secuencia na que non se dá coordinación pero que sería admisible:

Manuel Lourenzo: Témolle moito porque non coñezo outra Universidade, só coñezo esta, entón témolle bastante e creo que fundadamente. Agora, a pesar de todo creo que a aventura hai que correla e agora hai un único ámbito legal para organizar eses estamentos, e creo que merecería o esforzo de todo o mundo para facer chegar a eses estamentos unha mensaxe o máis colectiva posible e o máis intelixente posible que procedese de todas estas atrocidades que estamos dicindo aquí. [CORGA: TDN1997-03/2]

6. En estilo indirecto a conxunción *que* non pode preceder os marcadores que son conxuncións coordinantes pero si os que son adverbios:

É rico pero aforra moito.

*Dixo que era rico e engadiu que pero aforraba moito.

Dixo que era rico e engadiu que, non obstante, aforraba moito.

Aplicada esta proba ao noso problema, resulta o seguinte:

*Dixo que era rico e engadiu que agora aforraba moito.

*Dixo que era rico e engadiu que agora que aforraba moito.

?Dixo que era rico e engadiu que, agora ben, aforraba moito.

Acábensenos as probas que, en teoría, permiten a diferenciación entre unha categoría e a outra e seguimos sen chegar a unha decisión incuestionable. Parece que a gramaticalidade de *agora ben/ora ben* precedida de conxunción adversativa ou de *que* en estilo indirecto, así como a constatación do seu uso sempre entre pausas marcadas, chegando incluso a situarse en posición medial parentética, nos indica que debemos tratalas como locucións adverbiais, como así faremos no sistema XIADA.

Pola contra, case todas as probas apuntan á conxuntividade de *agora que/ora que*, excepto as pausas que a preceden e que habitualmente a seguen. Causanos bastantes problemas a duplicidade na escrita ante exemplos

a) con pausa posterior:

As trincheiras da guerra son un feito material, militar; ora que, por riba delas está o ideal dunha loita común pola mesma causa: a da legalidade democrática que se dera o pobo, refrendada polas urnas o 16 de febreiro de 1936, e que foi tronzada pola sublevación e a guerra. [CORGA: S. Álvarez, Memoria da guerrilla, Xerais, 1991]

_Eu non lle sei nada _repetiu ela, contundente, antes de virarlle as costas_. Aquilo pasou hai moito tempo. E non deron con el. Agora que, pregunte, ó mellor... [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

_Fixo ben. Fixo moi ben. Fixo moi ben. En canto á noite _mentiu, por non enredarse_, leveino ben. Agora que, se ma deixa aquí, agradeceríallo. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

b) sen pausa posterior:

Uns 400 voluntarios, capitaneados por Francisco Roca, coñecido líder intransixente de Ourense, ocupan as vilas de Tribes e Viana do Bolo, ora que a rápida e eficaz intervención do Exército e maila Garda civil aborta a sublevación. [CORGA: X. M. Moreno González, *O movemento obreiro en Galicia*, Xerais, 1990]

Se en vez de andar facendo preguntas por aí, viñese onda min directamente, tería adiantado moito. Máis do que cre. Agora que mellor sería terlle preguntado ó Xeneral, para non perder o tempo como o está perdendo nestas lérias. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Eu de ti, por moito ou por pouco medo que teñas, e perdoa por meterme onde non me chaman, deixaría todo como está. O pouco ben que podes obter nunca vai compensar o moito mal que se fixo e que se pode facer. Agora que alá ti. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

c) e incluso con pausa no interior:

_Home, haberaos ananos e de tódolos tamaños, supoño, eu que sei _dixo Ramiro_, ora, que os Mouros son unha cousa e os Ananos outra. [CORGA: X. Miranda, *Historia dun paraugas azul*, Xerais, 1991]

A cita vén tirada da revista *Muy interesante* _nº 275, abril de 2004_, e non sei que cousa será máis interesante, se a noticia en si, ou a propia constatación (para sorpresa de moitos e alboroto duns poucos) de que, efectivamente, os músicos si temos algo máis ca cerume entre orella e orella (agora, que o usemos para cousa boa, iso xa é outro cantar...). [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]

Por unha banda, nos casos de pausa interior, semella que o proceso de gramaticalización non acaba de reflectirse na escrita e que, debido á orixe da locución (*agora, que*), se pretende seguir mantendo no uso actual a grafía con coma intermedia coma se o *que* pertencese á cláusula que segue. Pola outra, a pausa posterior non sería óbice para considerar *agora que/ora que* locución conxuntiva, pois aínda que en xeral non se recomenda o uso de pausa gráfica tras conxunción, a non ser que vaia seguida por un inciso, non só é posible facer esa pausa senón que existen outras conxuncións que tamén o permiten, por exemplo a consecutiva *así que*:

Demasiados datos provocan situacións de colapso. Así que, que ocorre cando hai unha tarefa que absorve por completo a nosa atención? [CORGA: *CC2006-01/37*]

Non é humano ter que se erguer ás seis e media da mañá para ir traballar. Non é humano, non. Mais non lle hai que facer, e acabouse. Así que, aló foi para fóra da cama coma todas as mañás de todas as semanas de todos os meses de cada ano. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

De alí a pouco marchamos, así que, a distancia e o tempo disolveron aquel compromiso tan firme. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Así pois, o *agora que* adversativo será tratado en XIADA como locución conxuntiva.

Finalmente, quedánnos por expoñer as conclusións ás que chegamos para *agora/ora*, polo que imos con elas.

Se lembramos, os principais argumentos en contra da súa consideración como adverbios eran:

a) a posición que ocupan, e, en consecuencia, a falta de mobilidade.

A explicación a esta posición inicial pode residir no tipo de conector que é, xa que non atenúa, coma a maioría dos do seu grupo, senón que establece unha contraposición co que se dixo antes, un corte claro, que vai empecer ou poñer trabas á concreción do primeiro argumento; introduciría, pois, un argumento máis forte ca *non obstante*, *con todo*, etc., e neste sentido compórtase coma os conectores digresivos posto que dalgún xeito interrompe a información introducida.

e b), non documentarse precedido de conxunción adversativa.

Quizais precisamente pola contradición que implica a utilización dunha conxunción coma *pero* ou *mais*, as cales moitas veces son soamente conxuncións adversativas restritivas que atenúan mais non empecen, ao lado dun adverbio cuxa forza argumentativa é maior. Neste sentido, apóiánnos as palabras de Santos Ríó:

La proposición que *pero* presenta puede tener carácter adversativo restrictivo puro, es decir, neutral respecto de la favorabilidad o desfavorabilidad de unos hechos para con otros (*Los dos primeros son grandes pero los demás no*), en cuyo caso, la idea de contraste viene a veces reforzada por la presencia de alguna de las locuciones adverbiales deícticas de enlace interproposicional adversativo (en cambio, sin embargo, como contrapartida, por contra) colocada preferentemente en contacto, previo o posterior, con el elemento temático de la proposición, si lo hay, y siempre a modo de inciso parentético. Fumaba muchísimo, pero, en cambio, apenas probaba el alcohol. [Santos Ríó, 2003: 501]

Malia que o etiquetalos como conxuncións facilitaría a localización e extracción das ocorrencias nas que estas unidades presentasen valor adversativo, entendemos que categorialmente seguen sendo adverbios, aínda que cunha función discursiva e non sintáctica.

Figuran, xunto con *(ag)ora ben*, como conectores adverbiais contraargumentativos en Freixeiro (2005b), e esa mesma clasificación aparece para a variante completa *(ahora, ahora bien)* tamén en Martín Zorraquino e Portolés (1999: 4118-9) e Santos Ríó (2003: 182).

O feito de relacionarse semanticamente cun *antes* co que se contrapón (con este valor, nunca cun *despois*) fai que se producise a perda progresiva do seu valor continuativo e deíctico para destacar nestes casos só o adversativo. Así pois, prodúcese a gramaticalización e respectiva deslexicalización, ou como Vilela (1999: 267) a denomina, “dessemantización”, posto que *agora/ora* perdeu as súas calidades temporais funcionais destacando só a súa función conectora.

Polas características dos exemplos atopados, e só co ánimo de subliñar unha apreciación, *agora* é a variante dominante no rexistro oral (alocucións en teatro e entrevistas en ensaios, maioritariamente), mentres que *ora ben* predomina no escrito.

En resumo, con valor adversativo, *agora, ora, agora ben* e *ora ben* son categorialmente adverbios, mentres que *agora que* e *ora que* se encadran no grupo conxuntivo.

4.4.14.2.6. Cada vez que, unha vez que

Estamos afeitos a que os circunstanciais sexan desempeñados por adverbios ou frases adverbiais, pero existen tamén algunhas frases nominais que cumpren esta función. En xeral, sucede cos días da semana e con outros substantivos coma *vez, semana, día, mes, segundo, minuto, hora, véspera...*

Todos eles necesitan da presenza dun actualizador, ben o artigo, ben o demostrativo, ben un numeral ou ben un indefinido, aos que ademais pode sumarse un adxectivo, unha frase preposicional con *de* ou unha cláusula de relativo:

Os luns traballa na taberna, cando a muller do taberneiro pasa todo o día fóra, na cidade, coa irmá. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Este sábado termina a rodaxe da longametraxe televisiva *Lucrecia*, realizada por Mariano Barroso para Antena 3. [CORGA: *CG1995-02-03/175*]

Os leóns de San Mamés destrozaron o xardín de Vanderlei Luxemburgo este sábado pasado. [CORGA: *GH2005-02-22/51*]

Segundo fontes próximas á reunión, Chernomirdin viaxará de novo a Belgrado o luns que vén, na súa cuarta visita á capital iugoslava desde que o presidente ruso, Borís Eltsin, lle encargou mediar na crise de Kósovo. [CORGA: CG1999-05-22/24]

Fai tres noites, a víspera de estar aquí a última vez, non puido evitarme. [CORGA: L. Iglesias de Souza, *As alaudas do mencer. Pintura negra, en forma de folletín noveleiro, policíaco-humorístico-político-social e satírico, (tamén) en catro romances de cego, catro partes e certos comedios*, Edicións do Castro, 1989]

Vémolo tódolos días, a diferentes niveis (o apoio a Isaac Díaz Pardo, por exemplo). [CORGA: CC2006-06/48]

O obxecto desta epígrafe consistirá en delimitar categorialmente as combinacións *cada vez que* e *unha vez que*.

A presenza do substantivo *vez* en combinación co cuantificador de efecto multiplicativo *cada* e o *que*, que orixinariamente introduce unha cláusula de relativo, provoca que a construción amose significado temporal e se conmute usualmente por *cando*:

¿Acaso non percibes as miradas cargadas de xenreira que escapan dos seus ollos cada vez que che dá unha das miñas cartas? [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

"Cada vez que chove botámonos a tremer, e Urbaser levou todo antes de que chegasen as cámaras". [CORGAetq: GH2007-03-17/6]

Fíxoo cada vez que lle expulsaron a un xogador en once ocasións cada vez que lle pitaron un penalti en contra seis, cada vez que perdeu ou empatou en Balaídos once. [CORGAetq: GH2007-03-30/70]

Hoxe non podo evitar un arrepiño de culpa cada vez que lembro as treboadas de odio case criminal que aquela imaxe conectaba en min. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

De forma similar ao que sucedía con *sempre que*, e pensando no paralelismo de adverbio e frase nominal de carácter adverbial, atopámonos con dúas posibilidades de análise para a secuencia *cada vez que*: podemos optar por tratala como locución conxuntiva, tal e como se recolle nalgúns manuais do galego (por exemplo RAG/ILG, 2003 ou Álvarez e Xove, 2002), ou podemos decantarnos pola análise desagregada dos seus compoñentes, entendendo que estamos ante un indefinido, un substantivo feminino e unha cláusula de relativo que modifica á frase nominal que actúa como circunstancial de tempo.

Se a construción encabezada por esta expresión non é un sintagma deslexicalizado, o seu significado e función non pode variar con respecto a aquelas outras construcións nas que aparece illado e sen cláusula relativa que o modifique. Por outra parte, se a locución se gramaticalizou é previsible que presente unha forte cohesión entre os elementos que a integran.

A proba da coordinación demóstranos que *cada vez que* non constitúe unha locución fortemente cohesionada, por canto, malia que non é moi común, si se atopan no CORGA ocorrencias recentes de coordinacións do constituínte encabezado por *que*:

Antes de montar no auto fixen algo que acostumaba cada vez que ía á vila e que non quería que ninguén soubese, busquei a miña barca no porto. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]

España, engadiu, "rompía cada vez que unha muller non podía estudar, que era marxinalada ou que se vía sometida á dominación", e "o que fai forte un país son as leis de igualdade". [CORGA: GH2007-03-19/35]

Ó seu xuízo, "cada vez que Touriño conte unha mentira, que Quintana nos insulte ou que Pepiño saque a calculadora para dicir que malos resultados nos esperan, é que imos polo bo camiño". [CORGA: GH2007-03-26/9]

Polo que respecta ao emprego indistinto da frase nominal *cada vez* e mais da posible locución *cada vez que*, Pavón Lucero (1999: 640) afirma para o castelán:

Aunque en su formación no interviene ningún adverbio, podrían incluírse aquí las locuciones formadas a partir de algunos sintagmas nominales adverbiales. Entre ellas tenemos, en primer lugar, *cada vez que*. Este tipo de construcciones, sin embargo, sólo alternan con aquellas en que el sintagma *cada vez* no va seguido de una oración subordinada cuando en la principal hay una idea de gradación, como ocurre en (296a). Si en esta misma construcción *cada vez* aparece aislado y se suprime el adverbio *más* en la principal, el resultado, como refleja (296b), es de dudosa gramaticalidad. Lo mismo ocurre si en una oración como (296c), donde no hay gradación, se suprime la oración subordinada con *que*:

- (296) a. Cada vez (que lo veo) me parece más antipático.
 b. Cada vez ?(que lo veo) me parece antipático.
 c. Cada vez ??(que oigo decir algo así), no puedo evitar enfadarme

Decidimos comprobar se esta regra se lle pode aplicar ao galego, e para iso analizamos un pouco polo miúdo a distribución de *cada vez* na versión 1.5 do CORGA, onde si semella que non é tan indiferente a utilización dunha e doutra. Así, das 4170

ocorrencias que existen na versión 1.5 do CORGA coa frase exacta “cada vez”, se primeiro suprimimos aqueles casos nos que aparece seguida por unha cláusula de relativo, e a continuación imos prescindindo progresivamente dos diferentes elementos comparativos, observamos como se produce unha diminución considerable²⁰⁸:

-> NOT cada vez que:	3285
-> NOT cada vez m?is ²⁰⁹ :	1340
-> NOT cada vez menos:	1209
-> NOT cada vez m?llor:	1195
-> NOT cada vez p?or:	1174
-> NOT cada vez meirande:	1168
-> NOT cada vez maior:	1060
-> NOT cada vez menor:	1043

Ao acudir a ver os casos restantes, excluídos das condicións da busca todos os anteriores, atopámonos con que a tónica predominante é que na secuencia apareza ou ben algún elemento comparativo discontinuo de *cada vez* ou ben a frase nominal coa anteposición da preposición *de* (*de cada vez*), polo que decidimos engadir a expresión *de cada vez* á listaxe de elementos que queremos excluír da busca, quedándonos finalmente 980 ocorrencias para consultar.

Algúns exemplos de *cada vez* cun comparador presente, anteposto á expresión ou posposto, máis ou menos lonxe dela pero non ao lado, son os seguintes:

Entrevistado: Cada vez convénzome máis de que a postura inicial que adoptou Marrocos na sinatura do anterior acordo se está levando á práctica. [CORGA: CG2000-12-12/1]

Ata agora, os artefactos desactiváronse a tempo, pero é evidente que contan cunha infraestrutura suficiente como para enviar cada vez con maior frecuencia as súas 'mensaxes' explosivas. [CORGA: CG2000-12-19/54]

²⁰⁸ A busca realizada o 25/03/2012 sobre todo o corpus empregando o operador booleano NOT foi a seguinte: *cada vez NOT cada vez que NOT cada vez m?is NOT cada vez menos NOT cada vez m?llor NOT cada vez p?or NOT cada vez meirande NOT cada vez maior NOT cada vez menor*.

²⁰⁹ O uso do interrogante sérvenos de comodín para solventar os problemas que nos puidese ocasionar a variación acentual (*máis/mais*) ou a alternancia vocálica *e/i* (*mellor/millor; peor/pior*).

Fraga restoulles importancia ás declaracións feitas hai dous días polo conselleiro de Educación, quen bromeou sobre que a súa muller dixera que na súa casa cada vez se comería menos tenreira. [CORGA: CG2000-12-02/1]

Entrevistado: Pode haber pequenas matizacións pero en xeral todos foron alumnos moi bos, e cada vez con mellor nota. [CORGA: CG2000-11-12/50]

Mudou de médico, pero aquilo, cada vez, collía peor cara. [CORGA: A. Eiré, *Eu tamén fun coas vacas*, Espiral Maior, 2000]

Existen pequenos grupos que se defenden contra esa homoxeneización imposta, pero cada vez son menores e en moitos casos son perseguidos e aniquilados (os indios guatemaltecos) polo mero feito de existiren. [CORGA: L. Arias Rodríguez e R. Prieto-Puga Fariña, *Economía. Primeiro curso de Bacharelato*, Bahía, 1998]

Pero Siching xa non os acompañaba nas súas incursións, e as victorias eran poucas, e meirande cada vez o número dos que non volvían, caídos no combate. [CORGA: L. Fernández Marcos, *Galaxia lonxana*, Xerais, 1995]

Os meus pensamentos fanse máis lentos e máis torpes cada vez. [CORGA: L. Fernández Marcos, *Galaxia lonxana*, Xerais, 1995]

Antonia: Teño estado algunhas veces, menos cada vez. [CORGA: A. Avendaño, *Fin de acto*, Xerais, 1989]

Fóra os comparativos que xa vimos (*máis, menos, maior, menor, mellor, peor, meirande*), dáse tamén a concorrencia de *cada vez* con outras unidades para as que a comparación está implícita, por exemplo, cos adxectivos *superior, distinto, diferente*, ou cos adverbios *antes e despois*:

En calquera caso, a construción de unidades cada vez de superior tamaño podería ser explicada como o resultado da busca de economías de escala existentes aínda que non sexan decisivas e, simultaneamente, da elevación dos capitais precisos para entrar no ramo. [CORGA: X. Fernández Leiceaga, *Capital estranxeiro e industrialización en Galicia*, Xerais, 1993]

O noso patrón ordenou cortar os paus, e dous mariñeiros botáronse ó trafego, pero éralles difícil pois co movemento daban as machetadas onde cadrase, cada vez en lugar distinto, hastra que unha refolada adiantouse e arrincou dun saque a arboradura que saeu voando por derriba de nós. [CORGA: X. Neira Vilas, *Galegos no Golfo de México*, Edicións do Castro, 1980]

As guitarras acendían lumes nos corazóns anunciando un mencer cada vez diferente. [CORGA: M. Lourenzo González, *Arqueofaxia*, Xerais, 1995]

Por outra banda, estase "detectando que en Galicia cada vez se inician antes os nosos rapaces nos hábitos do tabaco e o alcol, e hai sobre todo un incremento entre as rapazas, que empezan a fumar e a beber a estas idades tan temperás", explica o delegado. [CORGA: GH2005-02-18/2]

sentábase, bebía un grolo mesto e dilatado, pousaba a xerra, cos movementos repetidos cada vez, e despois volvíase para deitar a vista distraidamente nos siareiros, tamén anónimos, daquela hora desusada. [CORGA: M. Lourenzo González, *Noites de papel*, Xerais, 1997]

El chegaba cada vez a unha hora precisa, tamén sempre a mesma. [CORGA: M. Lourenzo González, *Noites de papel*, Xerais, 1997]

Non obstante, fronte ao que se indica para o castelán na *Gramática descriptiva* (Pavón, 1999: 640), para o galego si se documentan casos de *cada vez* sen que teña que aparecer obrigatoriamente algún elemento de tipo comparativo. Os que seguen son só algúns deles:

Naturalmente que a chamara sempre que puido, polo menos un par de veces por semana, dándolle cada vez conta do resultado das cousas e prometendo tamén, de cada ocasión, que esta era a última chamada, que a próxima conversa ía ser cara a cara, labio a labio. [CORGA: X. R. Pena, *Paixóns privadas*, Xerais, 1991]

Era imposible dar con el porque o Lúnubro sabía cal era o sentido que o astrónomo utilizaba cada vez para tentar capturalo. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

Pero por fortuna, como levo dito, o meu xorne aberto e conciliador sóubose impoñer en todas as situacións e permitirme saír airoso e incluso enaltecido cada vez. [CORGA: M. Lourenzo González, *Noites de papel*, Xerais, 1997]

Entrevistado: Este quizais sexa un disco máis de escoitar, relaxadamente, cada vez vas atopando cousas novas, os timbres, os ritmos, etc. [CORGA: CG1999-05-13/60]

_O nome do outro _diciálle cada vez, e a negativa do roxo era o preludio a un novo premer no interruptor da corrente. [CORGA: L. Rei Núñez, *Expediente Artieda*, Xerais, 2000]

Escachou dúas ou tres ducias de ovos e cada vez afastaba a xema das claras antes de abocar aqueles ollos amarelos no ventre da amasadora eléctrica, que tiña un brazo que non tardou en atronar o local coas súas convulsións. [CORGA: L. Rei Núñez, *Expediente Artieda*, Xerais, 2000]

Estes profesionais non atopan diferencias entre nenos e nenas e din que cada vez redúcese a idade na que os cativos empezan a utilizar tacos e a falar de sexo. [CORGA: ANT2002-09-12/84]

O que me gusta da corredoira marítima de Rianxo é que está chea de parques (o da pasarela de madeira que desdobra a corredoira na Carballeira gústame especialmente), parques nos que

entras, cada vez con ollos novos (non están en folletos de axencia de viaxes) para mirar, descansar, falar, por falar, cos que sempre pasan á mesma hora. [CORGA: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Así pois, nin o criterio da coordinación nin o da funcionalidade nos indican que esteamos ante unha locución cohesionada. Agora ben, podemos intercalar entre a frase nominal e a cláusula de relativo algún elemento sen que se converta en agramatical? Na nosa opinión, o único caso que localizamos no CORGA é dubidoso:

San Lázaro puido asistir onte a unha completa coreografía: a dos xogadores do Compostela cada vez (e onte foron tres) que conseguiron bater a porta dun desafortunado Buyo. [CORGA: *CG1996-02-26/92*]

Dende logo, o que si non é gramatical é antepoñer unha preposición e converter o relativo en complexo (*o que*), dado que o antecedente, malia ser unha frase nominal, é funcionalmente unha frase nominal adverbial que, a semellanza dos adverbios para os que non é pertinente o xénero e número, non acepta a recuperación do xénero do substantivo *vez*:

O respecto que de sempre lle impuxera a figura do irmán primoxénito impeliáo a gardar silencio cada vez na que, a través do teléfono, intentara argumentarlle e este o interrompera, de xeito tan brután coma confiado e ledó, pra dicirlle que non habería problema ningún, que todo era normal e había ir avante do mesmo xeito. [CORGA: A. Conde, *Sempre me matan*, Ir Indo, 1995]

Finalmente, os mesmos criterios de restrición que guían o emprego de *cada vez* e *cada vez que* son extensibles ás estruturas paralelas formadas cos outros substantivos que tamén constitúen frases nominais de carácter adverbial: *ano*, *mes*, *día*, etc, e, non obstante, non se cuestiona se son locucións:

Cada ano que pasa estás máis guapa.

Cada ano estás máis guapa.

*Cada ano estás guapa.

En conclusión, no sistema de Xiada non tratamos *cada vez que* como locución conxuntiva, senón que consideramos tres unidades diferentes: *cada* (Idfs), *vez* (Scfs) e *que* (Tnaa). A caracterización do relativo, para o que non aplicará nin xénero nin número, será a que dará conta das peculiaridades adverbiais da frase nominal que constitúe o seu antecedente.

A diferenza do que sucede coa frase nominal *cada vez que*, *unha vez que* non amosa unha única análise. Por unha banda, está xa gramaticalizada para establecer unha relación de dependencia temporal entre dúas cláusulas; pola outra, segue a manter plenamente o seu valor léxico de tipo adverbial na análise desagregada. Vémolo claramente nos seguintes dous pares de exemplos:

a) locución conxuntiva:

Co paso do tempo e unha vez que tales políticas foron aplicadas neses países os resultados confirmaron algo elemental xa previsto polos críticos: [CORGAetq: ANT2002-11-07/29]

Concluído o ciclo, finalizado o movemento, detidos os pseudópodos, unha vez que foi expulsado o elemento estraño e rexeitable, todo volve ó seu ser. [CORGA: A. Conde, *Sempre me matan*, Ir Indo, 1995]

b) non locución:

Unha vez que lle tocara a esta de Colmeiro (que lla mercara a María Antonia, na Galería Torques; si, si, lembrábase ben), notou casualmente certa sensación difícil de explicar que non deixou de chafucalo, polo que, como empurrado por algo, seguiu a observar ata que comprobou debidamente que a engulemia lle desaparecía case ó instante en presenza da pintura. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

seica o contou moitas veces Zacharias, porque non se lle borrou da memoria que unha vez que foi tocar a Moscova viu que da cabeza dun home agromaba un arco da vella. [CORGA: I. López Silva, *Concubinas*, Xerais, 2002]

Os catro casos son conmutables por *cando*, mais iso débese á dupla faceta que caracteriza o adverbio relativo: a de desempeñar simultaneamente a función de elemento relacionante, nexo polo tanto, e a función de complemento circunstancial no interior da cláusula relativa. Agora ben, por un lado, nos casos de a) non podemos prescindir do segmento con *que*, mentres que nos de b) si é posible sen que se altere o significado; por outro, nos de a) non podemos conmutar o determinante por un numeral, por exemplo *dúas*, co conseguinte cambio de número no substantivo, mentres que esta conmutación é posible realizala nos de b), unha vez máis sen que se altere a significación da secuencia máis ca no paso dunha inconcreción (*nunha ocasión...*) á precisión numérica (*dúas veces...*).

Así pois, a pesar de estar ante as mesmas formas gráficas, a segmentación é distinta en función do papel que desempeña *unha vez que* en cada caso: nos de a),

locución conxuntiva de tipo subordinante [unha vez que Lcs unha vez que]; en b) determinante indefinido, seguido polo substantivo feminino *vez* e un relativo que introduce un modificador: [unha Difs un] + [vez Scfs vez]+ [que Tnfs que].

Unha vez que non demostra unha forte gramaticalización se temos en conta que é posible coordinar dúas cláusulas e que, nese caso, na segunda delas o coordinado só está encabezado por *que*:

Aos dezaseis anos casara coa infanta Blanca de Navarra, da que se separou trece anos despois, unha vez que se demostrou que o matrimonio non se tiña consumado e que a princesa seguía tan virxe coma antes da boda. [CORGA: X. Bernárdez Vilar, *Un home de Vilameán. Anatomía dunha revolución*, Patronato da Cultura Galega, 1976]

Grecia propón que unha vez que Belgrado acepte este acordo, e que os directores políticos do G-8 presenten un proxecto para a paz en Kósovo para ser aprobado no Consello de Seguridade da ONU, deberíase decretar unha tregua de 48 horas por tódalas partes implicadas na guerra (serbios, guerrilleiros kosovares e OTAN). [CORGA: *CG1999-05-21/18*]

Non obstante, a alternancia en igualdade de condicións de *unha vez que* e mais *unha vez* dános a entender que ambas presentan as mesmas características, e que a escolla depende soamente do falante, sen estar esta condicionada por ningún outro factor. É dicir, se comparamos os seguintes dous grupos de secuencias, os primeiros coa variante con *que* e os segundos sen ela, veremos que non existe ningún factor lingüístico –tempo, modo, etc.– que condicione a aparición dunha ou outra unidade:

a) unha vez que

Co paso do tempo e unha vez que tales políticas foron aplicadas neses países os resultados confirmaron algo elemental xa previsto polos críticos: [CORGA: *ANT2002-11-07/29*]

Juan José Lucas indicou a un grupo de periodistas que este sistema "só beneficia determinadas comunidades, pero asemade non queremos perder nin cinco céntimos, polo que, unha vez que deixamos patente cá l é a nosa postura, tratamos de que non se perda un peso para Castela e León". [CORGA: *CG1996-02-15/105*]

"O afán contributivo dos cidadáns crecerá unha vez que se rebaixen os impostos", engadiu. [CORGA: *CG1996-02-16/101*]

Cando están brandas, agrégase o peixe e, unha vez que este tamén coceu, escórrese case toda a auga. [CORGA: X. C. Rodríguez Sánchez e X. León Guillén, *Pesca submarina en Galicia*, Lea, 1997]

Concluído o ciclo, finalizado o movemento, detidos os pseudópodos, unha vez que foi expulsado o elemento estraño e rexeitable, todo volve ó seu ser. [CORGA: A. Conde, *Sempre me matan*, Ir Indo, 1995]

É toda aquela que é capaz de rebrotar unha vez que foi cortada. [CORGA: *F1997-06/3*]

b) unha vez

Unha vez estes ministros formen outra vez parte do Goberno, poderíase nomea-lo ministro principal. [CORGA: *CG2001-10-25/26*]

O maxistrado discrepante deberá "formular motivadamente o seu voto particular, que se incorporará á resolución da sentenza" que será notificada ás partes unha vez estea redactada. [CORGA: *CG2002-05-25/12*]

Por un lado, que o nacionalismo de CiU está pasado de moda, e polo outro, que a federación nacionalista abandonou un espacio político, o da "centralidade" que o PP debe ocupar unha vez se esgote o ciclo Pujol. [CORGA: *ANT2002-10-24/39*]

Roi Soares de Paiva, con alegría non finxida, unha vez ergueu o xoello, quixo acudir aos brazos do rei, como lle fora concedido en tantas outras ocasións. [CORGA: L. Villalta, *As chaves do tempo*, A Nosa Terra, 2001]

Unha vez foi atopado o cromosoma todo o demais tórnase superfluo e innecesario. [CORGA: D. Soutullo, *De Darwin ao ADN. Ensaíos sobre as implicacións sociais da Bioloxía*, Laiovento, 1997]

Entre os obxectos persoais que foron achados no camarote da señorita Águeda Charlotte Wien Hofmann o día 1 dos correntes, despois da nosa chegada ao porto de Recife, foi encontrada esta folla que polo seu interese particular lle remitimos unha vez nos foi imposible enviarlla á súa autora ao carecermos do seu enderezo postal. [CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]

Semella que a conxunción *que*, prescindible, se engade por analogía, por formar parte esta da maioría das locucións conxuntivas. En XIADA, ambas, *unha vez* e *unha vez que* son etiquetadas como locucións conxuntivas de tipo subordinante cando van seguidas de cláusula, excepto a de participio, coma nos exemplos de arriba.

Sen saber moi ben que condiciona a diferenza e, unha vez máis, a falta de descrições exhaustivas sobre o comportamento dos diferentes nexos e sen poder afondar no seu funcionamento, decidimos que no noso sistema *unha vez* en construcións de participio absoluto ou seguida por un adverbio ou frase preposicional que indique situación ou estado será catalogada como locución de tipo adverbial.

Polo que respecta ás construcións absolutas de participio, a supresión da locución non altera o significado, pois a indicación de que a situación descrita na construción absoluta é anterior á que se describe na cláusula principal achégaa xa o valor aspectual perfectivo do participio:

Vista por exemplo a posición actual da esquerda francesa, un non pode sustraerse á sensación de que entre Jospin e González hai un abismo. [CORGAetq: *TN1997-07/29*]

Vencido o primeiro contratempo, TomaCom non puido comercializar mexillón galego durante a recente marea vermella. [CORGAetq: *ANT2002-10-31/25*]

Superado o ecuador da lexislatura o Partido Popular está a mudar tanto as estratexias como as medidas políticas, económicas e sociais. [CORGAetq: *ANT2002-10-17/30*]

Sometidos a un cambio radical por unha forma aínda máis radical de televisión, os pacientes renacían da man duns maravillosos médicos que eran todo amabilidade e disposición _aí, nesa comparación, a espectadora perdía, acostumada como está a médicos sen tempo nin para darlle os bos días ao entrar na consulta_. [CORGAetq: *GH2007-03-30/85*]

No ano 96 foi relevado por Pérez Varela _á esquerda na foto_ e, chegadas as eleccións, foi excluído das listas do PP. [CORGAetq: *TN1998-04/12*]

A locución adverbial non fai senón reforzar a anterioridade e a perfectividade:

No entanto, unha vez chegadas a España, as mulleres eran distribuídas entre diferentes proxenetas. [CORGAetq: *DLV2007-05-15/81*]

Unha vez demostrada a validez do invento, a sociedade comezará coa busca dos materiais máis asequíbeis para comezar a fabricación seriada do produto, que se prevé iniciar cara finais do 2003. [CORGAetq: *ANT2002-09-12/29*]

Unha vez descontados os beneficios atribuídos ós minoritarios, o beneficio neto quedou en 84.013 millóns, o que implica un crecemento do 16,2 por cento. [CORGAetq: *CG1996-02-25/77*]

O malo é se, unha vez conseguida, se chega a perder; [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Os nosos maiores, unha vez comidos e bebidos, saían a refrescar a quentura das caras bailando unhas pezas, que alá executaba unha orquestra de cinco profesores, cinco, vestidos con traxes color crema. [CORGAetq: X. M. G. Trigo, *Extintos básicos*, Xerais, 1998]

Ese mesmo reforzo aparece tamén con outros adverbios temporais coma *despois* ou *xa*:

E o caso é que só despois de pasadas moitas décadas as innovacións citadas xeraron unha vaga de prosperidade xeral, comenta o profesor. [CORGAetq: *CG1996-02-24/1*]

Nas fotos que acompañan esta reportaxe apréciase o estado no que quedou o territorio despois de feita a concentración. [CORGA: *GII1996-03/49*]

Despois de feito o corte collemos o mazo coas figuras cara abaixo e imos quitando unha a unha pola parte de arriba e virándoas segundo as imos quitando. . [CORGA: M. Aneiros, *Da antiga arte de bota-las cartas. A baralla española*, Vía Láctea, 1991]

Sen dicir palabra durante algúns minutos, alonxáronse de toda zona transitada e, xa convencidos de que ningún habitante do labirinto podía escoitalos, preguntáronse ó mesmo tempo: [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Xa morto mi padre, a mamá dícialle ao teu papá, Eduardiño, mira ben que túa irmá e mais eu somos a familia que che queda. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

E xa que se prevén numerosas tandas de visitantes nos meses vindeiros, a modo de guía e sobre todo para os non iniciados nestas artes, aconsello especialmente os lavadoiros (onde as mozas quedan lavando), as airas vellas, onde as avoas xogan a brisca co pano desatado, e xa perdido o medo, canella adentro, os corredores altos e as solainas onde mulleres grosas e fartas crian as camadas entre cordeis de roupa e fatelos brancos. [CORGAetq: C. Pereiras, *pequenos infinitos*, Difusora de Letras Artes e Ideas S.L., 2008]

Precisaba botarse a durmir e íao facer non ben pasada Palencia, aínda que só fose durante un cuarto de hora. [CORGA: F. Alonso, *Trailer*, Xerais, 1991]

Finalmente, no concernente a *unha vez* mais estrutura locativa, non podemos pensar en locución conxuntiva sen recorrer á elipse verbal, por exemplo do verbo *estar*, e tampouco aquí pode alternar con *unha vez que*, polo que cremos que estamos ante unha unidade de tipo adverbial. Vexamos algúns exemplos:

Unha vez baixo o chorro, mergullados nunha néboa mesta, ensarillaron unha conversa de parella adulta e demorada, con tranquilidade no agarimo e poucas urxencias, coma se estivesen certos da continuidade da noite, coma se soubesen que non había... [CORGA: X. Queipo, *Papaventos*, Xerais, 2001]

Unha vez arriba, corremos cara á rúa coma animas escapadas do inferno. [CORGA: X. A. Neira Cruz, *As cousas claras*, Xerais, 2001]

Unha vez en vea comunicativa, Carlos non se detiña. [CORGA: Ú. Heinze, *Culpable de asasinato*, Xerais, 1993]

Unha vez dentro, esfregou unhas cuncas e pousounas na mesa. [CORGA: L. Villalta, *As chaves do tempo*, A Nosa Terra, 2001]

Aquel grupo, a cada intre reunía máis xente de ambas cores e Catuxo viuse arrastrado pola multitude que se dirixía de novo ó hipódromo onde, segundo o programa, debían ter lugar os xogos do segundo día, pero unha vez alí, cando o emperador volveu a facer acto de presenza, os berros contra el renováronse esixíndolle a liberación dos reos. [CORGA: F. A. Vidal, *A derradeira decisión*, Toxosoutos, 2001]

Unha vez no exterior a súa euforia cantora non minguaba polo que, durante un par de minutos a noite non fixo máis que devolver para dentro os ecos dunha canción de esmorga que empezaba: [CORGA: V. Araguas, *A canción do verán*, Ir Indo, 2001]

Unha vez no coche volveu esperar inutilmente a explicación que non recibira pola noite, nin tampouco na hora do almorzo. [CORGA: A. Conde, *A casa de Adara*, Xerais, 1996]

Non, se eu unha vez no prato si que as como. [CORGA: M. Aleixandre, *Lobos nas illas*, Xerais, 1996]

Unha vez fóra do garaxe, Fátima preguntou cara onde collía. [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]

A locución adverbial non é suprimible sen que se altere o significado ou a secuencia deveña en agramatical e é conmutable polo adverbio de natureza perfectiva *xa*:

? Dentro, esfregou unhas cuncas e pousounas na mesa.

Xa dentro, esfregou unhas cuncas e pousounas na mesa.

* En vea comunicativa, Carlos non se detiña.

Xa en vea comunicativa, Carlos non se detiña.

En resumo, *unha vez* ~ *unha vez que* mais cláusula son tratadas en XIADA como unidades conxuntivas, mentres que *unha vez* en construcións de participio absoluto ou estruturas locativas é considerada unha locución adverbial.

4.4.14.2.7. Nin que, aínda que, mesmo que, por moito que, por máis que, por+adx+que

Naturalmente, non imos centrarnos na combinación de conxunción coordinante *nin* seguida de cláusula con *que*, exemplos 1 e 2, nin tampouco na combinación casual do adverbio *nin* seguido pola conxunción *que*, exemplos 3 e 4, senón na combinación habitualmente clasificada entre as locucións concesivas (Álvarez e Xove, 2002: 164; RAG/ILG, 2003: 184), exemplos 5 e 6:

- 1) Desde logo, non se conseguirá con isto que todos os traballadores europeos gañen igual nin que se produza unha igualación de custos e prezos; [CORGAetq: *TN1998-05/18*]
- 2) Non fai falla dicir de onde ven Silvester Stallone, nin que Jean Claude Camille François Van Varenberg, alias "Van Damme", é holandés. [CORGAetq: *ANT2008-07-10/51*]
- 3) Non hai nin que dicir que nunca puiden voar por semellantes sistemas. [CORGA: M. Murado, *Memoria de derribos*, Galaxia, 1993]
- 4) _Aí estás, querería matarme antes que seguir mirando para esa cara, de feito non creo nin que sexa quen de armar unha nova frase sen tatexos. [CORGA: M. Regueira, *L'affiche rouge*, Xerais, 2007].
- 5) ¡Non cachan unha nin que os maten! [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]
- 6) Sancha: Ti non te ordenas monxa nin que veña o bispo mesmo a te bendicir. [CORGA: T. Moure, *Unha primavera para Aldara*, Deputación Provincial da Coruña, 2008]

Nun primeiro momento pensamos que tamén a combinación co pretérito de subxuntivo encaixaba nesa significación (exemplos 7 e 8):

- 7) _¿Pero de que cona te ris, Helena? Nin que foses ti a que te metes cousas. [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]
- 8) "Eu, señor, xa de naipelo me sentía como incómodo dentro do meu peleslo, que nin que soubese que non era aquel o meu sitio na vida. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Así é que, cando nos cuestionamos se estabamos en realidade ante unha locución conxuntiva ou se, polo contrario, nos atopabamos con dous elementos independentes, incluímos no mesmo paquete as dúas especializacións de significado (DRAG, 1997):

1. Introduce unha circunstancia que, se fose real, podería facer comprensible, factible, etc., o que se mencionou, propuxo, etc., anteriormente. *¡Nin que foses un neno!*
2. *loc. conx.* Introduce unha frase que dificulta pero non impide o cumprimento da anterior. *Non como caracois nin que me paguen.*

Os tres argumentos que nos van servir para decidir se a tratamos no sistema como locución conxuntiva subordinante son os seguintes:

- 1) Rección modal. Fronte ás outras combinacións de *nin* e *que*, que poden aparecer con indicativo ou subxuntivo, a suposta locución rexe subxuntivo (exemplos 5, 6, 7 e 8). Xeralmente constrúese co pretérito (acepción 1 do

DRAG, exemplos 7 e 8), mais tamén é habitual co presente (acepción 2, exemplos, 5 e 6).

2) O significado achegado pola posible locución, expresión de forma hipotética~irreal de algo cuxa realidade se nega enfaticamente, non é deducible da suma dos seus compoñentes.

3) Gramaticalización ou cohesión interna da locución. Aquí é onde empeza a quebrarse a unicidade da combinación. En principio, na locución non se podería introducir ningún elemento entre os dous constituíntes, pois a gramaticalización debe ser completa para a súa consideración; non obstante, esa forte cohesión que si se dá para *nin que* na acepción 1 do DRAG, a única que nós imos considerar locución conxuntiva, non ten lugar coas outras combinacións de *nin e que* (conxunción coordinante + conxunción introdutora de subordinada substantiva - exemplos 9 e 10- e adverbio + conxunción -exemplo 11-) onde é posible a intercalación, por exemplo, dun adverbio:

- 9) Pero a crise do sindicato non proba a diminución da solidariedade, nin tampouco que desaparecesen as tensións propias do mundo do traballo. [CORGA: CG1999-05-02/1]
- 10) É verdade que a el non lle gustaba que ela carecese de sentimentos de temor, por máis que nunca falaron da historia esa do velorio en solitario cun cadáver, nin que manifestase a disposición a experimentar calquera sensación nova, nin moito menos que expresase desexos de saltar en paracaídas. [CORGA: C. Casares, *Deus sentado nun sillón azul*, Galaxia, 1997]
- 11) Porque o azar era impredecible e non excluía nada. Nin sequera que saíse moi axiña daquel inferno. [CORGA: C.G. Reigosa, *Narcos*, Xerais, 2001]

Tampouco a suposta locución da acepción 2 do DRAG é allea á inserción de algún elemento entre o *nin* e o *que*, só que neste caso os elementos intercalados son membros do paradigma de estruturas que habilita a lingua para expresar a concesividade: *anque, mesmo, aínda, por moito...* (exemplos 12, 13, 14 e 15):

- 12) Eu, poño por caso, hai unha comida que non a probaréi nin anque me poñan unha pistola no peito. [CORGA: C. Gonsar, *Arredor do non*, Edicións do Cumio, 1995]
- 13) Alcenderon os pitos, botando o fumo abaixo, que non se podía perder, nin mesmo que fose unha mistura de follas de pataca. [CORGA: E. Alonso, *A rúa do noxo*, Galaxia, 1977]

- 14) _Contigo e co De la Huerta eu non me meto a nada, nin aínda que soubese que tiñades as chaves da Gloria. [CORGA: X. Alcalá, *Código morse*, Xerais, 1996]
- 15) Os dedos das mans non bastaban para colle-los peixes nos ríos sobre tres palmos de auga, nin por moito que se esforzasen os homes dos primeiros albores da humanidade. [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

Xa Álvarez e Xove (2002: 164-165) inciden en que nalgúns “*dos sintagmas anteriores pódese omitir que, resultando unha preposición ou locución prepositiva (malia, a pesar de, pese a) como nexo da oración concesiva, con verbo en Inf. (...) Tamén os adverbios inclusivos forman parte de sintagmas concesivos (aínda, mesmo e nin, aos que se une incluso), teñen ese valor seguidos de oración de Xer.*” para continuar afirmando que “*estes mesmos adverbios poden reforza-las oracións de Inf. e as modotemporais*”.

Da primeira afirmación xa tratamos na epígrafe 1 deste apartado referido a locucións conxuntivas. Da segunda interésanos concluír que:

1) O feito de existiren exemplos abondos cos adverbios inclusivos mais xerundio, dos que recolleemos deseguido unha pequena mostra, proba que cando van en combinación con *que*, malia a súa significación concesiva, non forman locución conxuntiva:

- 16) I eisi, "Cantares gallegos", aínda sendo un libro auténtico, non acada as derradeiras raigañas do propio ser rosalián. [CORGA: D. García-Sabell, *As facianas do erotismo contemporáneo*, Galaxia, 1976]
- 17) Xa ven, é unha historia tan frecuente tan de chorar que podería dar noxo reality-show barateiro, pero é miña dalgún xeito, é o que me deixaron os meus pais, porque eu son xa unha muller forte que non chora nin pelando cebolas. [CORGA: M. Seixas, *Viñeron do espacio interior*, Edicións Positivas, 1996]
- 18) Hai xente que, mesmo sendo boas persoas, non entenden nada. [CORGA: D. Ameixeiras, *Tres segundos de memoria*, Galaxia, 2007]

2) O feito de que todos eles, excepto loxicamente *aínda* por estar xa amalgamado na conxunción, son compatibles coa conxunción concesiva coloquial *anque*²¹⁰ reafirma a non consideración como locución:

- 19) Aínda que as investigacións non fixeron máis que principiar, todo apunta a un roubo perfectamente deseñado por unha organización profesional ó servizo de traficantes de arte internacional que leva operando..." _¡Traficantes internacionais de arte! ¡Non cachan unha nin que os maten! ¡Ha, ha! ¡Internacionais! [CORGA: Jaureguizar, *Comendo espaguetis diante da televisión*, Xerais, 1998]
- 20) Todo o día mirando para o chan. E non se che levanta nin anque chames unha grúa. ¡Ca! Nin anque lle poñas música. [CORGA: A. R. Ballesteros, *Xogos de damas*, Xerais, 1988]
- 21) O crepúsculo de Natal abateuse rápido sobre os Arcos, con calixines de neve e de tristura. Meu padriño, mesmo que o cuarto estivese case en sombras, seguía a ler nos folios amarelecidos, ou se cadra recitaba de cor un texto mil veces repasado. [CORGA: X. L. Méndez Ferrín, *Arraianos*, Xerais, 1991]
- 22) "Na outra fronte", engade, "están os latifundistas e a alta burguesía en xeral, que non admiten cambios mesmo anque sexan graduais e parciais, da orde social existente e móvense para que o exército federal aniquile o Exército Zapatista de Liberación". [CORGA: *CG1995-02-23/59*]

3) O feito de que a locución conxuntiva subordinante concesiva por excelencia, *aínda que/inda que*, na súa variante normativa, e a conxunción *anque*, na non

²¹⁰ Eliminada da nómina de conxuncións concesivas na última reforma das normas ortográficas do galego (RAG/ILG, 2003), non deixa de ser unha cuestión controvertida. Ferreiro (1999: 372), así como a mesma Academia galega (RAG/ILG, 1995: 183), destacan que malia a súa semellanza coa conxunción castelá non pode rexeitarse a hipótese dunha redución popular. Con todo, non é verdade que non se documente ata Sarmiento, pois no TMILG (*Tesouro Medieval Informatizado da Lingua Galega*), [consultado o 21/02/2012], documéntanse 3 ocorrencias: unha do século XIII e dúas do XV. A do XIII (documento notarial de Zamora), na nosa opinión, é castelán e reproduce polo contexto o segmento "a uno que", carecendo de calquera sentido concesivo ou adversativo. As outras dúas, malia que hai que tomalas con moita reserva, pois non é o primeiro caso en que as lecturas non reproducen os manuscritos, parecen válidas:

1	1292	VIM	18	Et porque sobr -esto veo	anque	Dom Alvaro, vuestro Bispo, me mostrou	Orde
2	1475	SDV	83	suçesores no se divida por herdeyros,	anque	sejan mortos, salvo que sempre jamays	Foro
3	1492	SMCP	35	contrabto se pode e deva desatar	anque	por nosa parte seran ditas e	Testamento

normativa, poidan ir acompañadas de xerundio, ratifica que só *aínda/inda que* é unha verdadeira locución:

- 23) Dendes daquela non voltei velo diante, anque andando un bo tempo, na primeira estada que botei no hospital, ó saberen que eu era sobriño seu, os médicos quixérono avisar, pero alporiceime de tal xeito con eles que acabaron por desistir da idea. [CORGA: X. M. Martínez Oca, *Número de patente / Querido Brais / A figueira*, Edicións do Castro, 1978]
- 24) Inda que atendendo á definición da gramática, "disciplina científica que se ocupa das funcións da palabra no interior do discurso", non é doado deducir que en realidade é un nexa entre a fonoloxía e a semántica, entre a música das palabras e o seu significado. [CORGA: CG2001-10-06/16]
- 25) Dese xeito, aínda que baixando algo respecto a datas anteriores, esa oferta de Antena 3 foi a raiña da programación de noite no top da audiencia no eido estatal. [CORGA: GH2007-03-02/96]

Nos demais casos atopámonos cun adverbio (*nin, mesmo, incluso*) xunto co cal a conxunción *que* amosa valor concesivo²¹¹; mais na nosa opinión non conforman locucións. Así pois, de todas as combinacións que vimos de ver nesta epígrafe soamente *nin que* seguida de pretérito de subxuntivo na súa significación modal condicional irreal e mais *aínda/inda que* van ser para nós locucións conxuntivas.

Non queremos rematar este apartado sen tratar, xa que mentamos a expresión da concesividade, as combinacións *por máis que, por moito que, por + adxectivo + que*, etc. Algúns exemplos podemos velos en:

Por moita retórica que lle poñan, non é fácil ocultar que desde mediados dos 70, agás no bienio 86-88, Galicia sofre unha continuada caída do emprego _141 mil nos últimos sete anos_. [CORGAetq: TN1997-05/1]

Aínda hai territorios virxinais nalgures, por estraño que pareza. [CORGAetq: GH2007-03-17/85]

²¹¹ Non o documentamos na gramática de Ferreiro (1999), mais tampouco este autor realiza un estudo detallado dos múltiples valores da conxunción multifuncional *que*. Si o recolle Freixeiro (2005b: 240) a partir da listaxe tirada de Huber para os conectores concesivos na lingua antiga. Tamén chegou ás nosas mans, por casualidade, outro estudo sobre a expresión da contraargumentación, este no castelán medieval, no que se documenta este uso e que nos serve, xunto co de Freixeiro, para apoiar a nosa tese:

Dígote, pues, que sy tú que amas jamás levantaste falso testimonio contra alguna o alguno por amor de aquella que amavas. Que digas non, yo te lo pruevo. (Corbacho 98: 5) [Traballo de Investigación Titorado de C. Manzano Rovira: La contraargumentación oracional y discursiva en el castellano medieval, Departamento de Lingua Española, USC, 2005/06]

Grazas, Carme, pola túa xenerosidade ao cederme o acceso a este e outros traballos e por soportar, mesmo participar gustosamente, os meus desvaríos locucionais.

E é que os neurocientíficos Marc Hauser e Antonio Damasio acaban de contar en Nature que, se a unha persoa co córtex prefrontal ventromedial danado, como Gage, un lle pregunta se mataría o seu fillo para salvar dez, dez fillos dalgúns outros, a tal persoa diríalle que si, por pura matemática, dez a un, e por moi repugnante que poida parecer o cálculo, cando vostede e mais eu salvaríamos o fillo, aínda a conta de condenar a dez e de mil malas noites, por moi repugnante que poida parecer o cálculo, neste caso tamén. [CORGAetq: GH2007-03-30/90]

Rueda advirte que non cesarán "por moito que el perda por nervios" [CORGAetq: GH2007-03-17/22]

pero agora xa non son unha forza, por máis que poidan facer aínda algunha que outra barbaridade. [CORGAetq: ANT2008-07-10/84]

En ningún momento cuestionamos se calquera das expresións anteriores introduce a ANTÍTESE dunha estrutura concesiva²¹², non obstante si cremos que isto por si só non debe conducir a que as unidades que as introducen sexan incluídas na nómina de locucións conxuntivas.

En xeral, na tradición portuguesa, igual ca na galega (Costa Casas et al., 1988; RAG/ILG, 1995 e 2003; Álvarez, Regueira e Monteagudo, 1986; Freixeiro, 2000; Álvarez e Xove, 2002 e Hermida, 2004), é habitual que *por mais que* e *por muito que* engorden a listaxe de conxuncións concesivas; así recóllenas, por exemplo, Cunha e Cintra (2005: 582) ou Vilela (1999: 260). Non obstante, existen tamén outros manuais nos que se profundiza un pouco máis, e xa non tratan as estruturas anteriores como locucións conxuntivas, senón que as integran noutros medios lingüísticos que expresan concesividade. É o caso de Mira Mateus et al. (2003: 720-721):

Outra forma que associa valores concessivos e comparativos são as orações concessivas intensivas, em que se põe em contraste a intensidade de uma qualidade ou a quantidade de uma substância e o conteúdo proposicional expresso pela oração principal; na oração concessiva é selecionado sempre o conjuntivo:

Por muito trabalhador que ele seja, não conseguirá o emprego.

Por mais que ele seja trabalhador, não conseguirá o emprego.

Por muita comida que ela coma, não engorda.

Por muito bem que eles cantem, não gosto de os ouvir

Uma variante da construção anterior são orações que articulam nexos relativos e concessivos, como nos exemplos seguintes:

²¹² Para máis detalles véxase, por exemplo, Rojo (1978).

A rapariga, inteligente que seja, vai ter dificultades porque trabalha pouco.

Mil felicitações que eles me mandassem, eu não esqueceria o que me fizeram.

Os problemas, muitos que fossem, não impediram a continuação do rali.

Na reflexión que segue, acerca da pertinencia de valorar se estamos ante locucións conxuntivas ou non, imos deixar xa fóra a construción con adxectivo *–por + adxectivo + que–*, á que nos permitimos agregar tamén a de con adverbio, *–por + adverbio + que–*, malia non recollerse esta última nos distintos manuais, debido en ambos os casos á súa indiscutible mutabilidade e ao seu incontrovertible funcionamento sintáctico no interior da cláusula de relativo que segue. Un par de exemplos ilustrativos das construcións con adverbio ou adxectivo constitúeno os seguintes:

Compostela debe cooperar con todas, por lonxe que estean e por moi insólitas que sexan. [CORGA: CG1996-02-18/1]

facía todo o que a tía lle encargaba sen torcer o xesto nin protestar, por moi lonxe que o mandase. [CORGA: U. Casal, *Vidas exemplares*, Xerais, 2006]

De saber que o seu destino inmediato era a morte, por moi complexo ou ilegal que fose o asunto que se traía entre mans, só un suicida se negaría a toda axuda. [CORGA: M. Forcadela, *Barato, barato*, Sotelo Blanco, 1993]

Por moi ben que a tratase, por moitos bombóns, entradas e pendentos que houbera, íalle pedir novamente os cartos por última vez. [CORGA: X. R. Pena, *Paixóns privadas*, Xerais, 1991]

Dende estas alturas, nas que as campás de Trascorgos se senten tan ben _e a quen non lle gusta ter novas da casa, por moi a gusto que se atope na capital_, o Marqués de Escandoi _López de Somoza de apelido primeiro, que ninguén se acorda_ ve a miúdo cómo Guillermo, antano tan orgulloso, deambula coa cabeza afundida entre os ombros. [CORGA: R. Loureiro, *O corazón portugués*, Galaxia, 2000]

Non acostuma recollerse tampouco nos manuais, mais debería sumarse así mesmo ás estruturas das que falamos, a que conforma *por pouco que* e as súas variantes:

Mais Ramón non lle vía un temperamento bastante enérxico para chegar a afirmar unha vontade de heroe, por pouco que fose. [CORGA: A. Risco, *Mascarada*, Galaxia, 1995]

Mackie: Por pouca cousa que fose, dime que era. [CORGA: R. Vidal Bolaño, *A ópera de a patacón. Versión libre para charanga e comediantes pouco ou nada subsidiados, arredor de textos de John Gay e de Bertolt Brecht*, Xerais, 1998]

Comprenderás, por pouco intelixente que sexas, que se eu, que teño a obriga divina de ostentar o poder, non me sinto autenticamente comprendido, non podo, e non debo deixarche responder... non podo permitirme e permitirche que perdamos o tempo con palabras baleiras. [CORGA: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]

Eu sempre teño que sacar algo de agasallo por moi pouca utilidade que poida ter, ¡ha, ha! [CORGA: M. Seixas, *A velocidade do frío*, Xerais, 1996]

Tanto a proba da coordinación como a da invariabilidade responden negativamente á posibilidade de tratar *por máis ~ máis(...) que, por moito ~ moito/a(s) (...) que e por pouco ~ pouco/a(s) (...) que* como locucións conxuntivas:

E a ninguén lle permitía que se achegase a ela máis da conta, por máis que lles dixese que cada cousa ó seu tempo, e que logo habían facerlle isto e o outro e o de máis alá. [CORGA: X. Fernández Ferreiro, *Agosto do 36*, Xerais, 1993]

Entrevistado: A ver, eu vivo no medio rural, e por moito que un partido faga unha campaña determinada e que conte co apoio dos medios para maquillar os erros, tecer ou terxiversar unha realidade, decátaste do que pasa, non ten que contarcho ninguén. [CORGA: *TN1997-11/11*]

¡Ai, Xesús!, nin por moitos cursiños que fixera, comentaba a recelosa nai, non lle deixo eu as filliñas nas mans dun home por nada do mundo, nin siquera do meu. [CORGA: M^a X. Queizán, *A semellanza*, Sotelo Blanco, 1992]

Sempre fora así e, por máis monumentos que o tempo erguese en Compostela, non ía ser de diferente modo nesta ocasión. [CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

Eu de ti, por moito ou por pouco medo que teñas, e perdoa por meterme onde non me chaman, deixaría todo como está. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Uns materiais que nunca poderán seren substituídos por moita, xenerosa e inmediata que sexa a indemnización que dean as institucións. [CORGA: *ANT2003-01-30/85*]

Unha mañá, que polo demais amenceu boa, Nanook deixou de respirar, e por moitas e moi desesperadas manobras que tentou primeiro Mirabeau, con máis ansia que coñecemento técnico, e logo o cirurxián, con máis técnica ca empeño, non volveu respirar máis nunca. [CORGA: X. Queipo, *O paso do noroeste*, Sotelo Blanco, 1996]

Ceferino: Hai que se foder, por moitos nomes de grupos en inglés que aprendades a pronunciar seguiredes sendo de Vallecas ou de Teis ou de Urzaiz, estades condenados a dicir reseso, esmagar ou colo, ignorando que son palabras galegas; [CORGA: A. A. Vila e X. Cid Cabido, *Copenhague. Comedia evidencialista*, Xerais, 1993]

Ao anterior temos que sumar que hai evidencias de que a partícula que segue ao grupo cuantificador non é conxunción senón pronome relativo, tal e como demostra a

anteposición das frases preposicionais seguintes, onde se ve claramente que os grupos preposicionais “*con que*” e “*das que*” se constrúen respectivamente cos predicados *bater* e *axustar* e mais co atributo *cheo*.

Machetei e machetei, pero aquilo máis que pau semellaba ferro e o fio da macheta non pasaba a tona por máis forza con que lle batera. [CORGA: X. M. Martínez Oca, *Número de patente / Querido Brais / A figueira*, Edicións do Castro, 1978]

Por moito tino con que axustásemos o cálculo e trazado das posíbeis traxectorias, poñendo en xogo brazos e varas tendidas diante dos ollos en horizontal guieira, apuntando ao lonxe nas direccións millor enfiadas estimando a incidencia dos distintos terreos, obstáculos en que bater, correccións previstas... o resultado era sempre o mesmo: [CORGA: M. Riveiro Loureiro, *Premios Pedrón de Ouro 1985/1986*, Edicións do Castro, 1987]

Hai anos que, xunto con outra moita xente, teño escrito que un acto terrorista é unha declaración de guerra se se comete contra un estado democrático de dereito, por moi cheo de eibas das que este estea cheo; [CORGA: *CG2001-10-15/19*]

Semella que todas estas estruturas de sentido concesivo son variables e conforman un grupo cuantificador que parece desempeñar unha función sintáctica no interior da cláusula que o acolle. Isto é ben evidente nos seguintes exemplos, onde podemos observar non só a cuantificación senón tamén a inclusión de modificadores de distinto tipo no interior do grupo cuantificador, e mesmo no último deles coordinación entre os adxectivos:

E, por máis vontade que poñía, desde o mesmo momento en que subía ó aparato, na súa cabeza aloxábase unha conclusión trascendente e desencantada: [CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

Por moitas pescudas intuitivas que o poeta realice, sempre se lle escapará, esguía, a derradeira esencia de ambas instancias. [CORGA: D. García-Sabell, *As facianas do erotismo contemporáneo*, Galaxia, 1976]

A Comunidade Autónoma pode acudir a tódalas fontes que considere oportunas, por moi lonxanas e alleas que sexan ó seu propio espazo xurídico. [CORGA: B. Cores Trasmonte, *Dereito autonómico de Galicia*, Xerais, 1987]

Se realizáramos a análise sintáctica, veríámonos na obriga de asignar as funcións de Complemento Directo nos dous primeiros exemplos anteriores (*por máis vontade que poñía*; *por moitas pescudas intuitivas que o poeta realice*) e Atributo no terceiro (*por moi lonxanas e alleas que sexan*).

Estas construcións caracterízanse por estaren os grupos cuantificadores sempre antepostos e situárense entre a preposición e o elemento de enlace. A anteposición das expresións cuantificativas expostas aos predicados cos que se constrúen garda relación, en certa medida, coas construcións enfáticas con artigo:

É un feito probado o facilmente recoñecibles que resultan os técnicos de son no contorno dun concerto. [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]

Apoia esta relación o feito de que o grupo sintáctico anteposto poida segregarse, de forma que só se antepoña un dos seus fragmentos:

Por moitas ganas que teña de contempla-los cadros, ten que esperar a que o invite para admiralos, e sexa discreto no seu xuízo co fin de recibir cantas explicacións precise. [CORGA: L. Prego Carregal, *Saber ser, saber estar. Normas-Consellos-Curiosidades*, Toxosoutos, 1994]

Non imaxinaba un casamento a base dun engano por moi adicta que fose ás mentiras. [CORGA: Ú. Heinze, *Culpable de asasinato*, Xerais, 1993]

Así pois, parece claro que nas construcións que inclúen frase nominal ou substantiva -non é relevante aquí a distinción terminolóxica- entre a preposición e o *que*, en moitas ocasións coa inclusión de modificadores e mesmo coordinación entre os modificadores, o grupo cuantificador desempeña unha función sintáctica no interior da cláusula que o acolle. Nestes casos, entón, non estamos ante locucións e, en consecuencia, no sistema de XIADA non seguimos para eles a tradición de análise existente, senón que procedemos a desagregar os distintos compoñentes. Por exemplo:

- a) por máis vontade que poñía: P [Preposición]; Wm [Adverbio, modificador]; Scfs [Substantivo, común, feminino, singular]; Tnfs [Relativo, non determinante, feminino, singular]; Vii30s [Verbo, copretérito, indicativo, 3ª, xénero non aplica, singular]
- b) por moitas pescudas intuitivas que o poeta realice: P [Preposición]; Idfp [Indefinido, determinante, feminino, plural]; Scfp [Substantivo, común, feminino, plural]; A0fp [Adxectivo, grao non aplica, feminino, plural]; Tnfp [Relativo, non determinante, feminino, plural]; Ddms [Artigo, determinado, masculino, singular]; Scms [Substantivo, común, masculino, singular]; Vps30s [Verbo, presente, subxuntivo, 3ª, xénero non aplica, singular]
- c) por moi lonxanas e alleas que sexan: P [Preposición]; Wm [Adverbio, modificador]; A0fp [Adxectivo, grao non aplica, feminino, plural]; Cc [Conxunción, coordinante];

A0fp [Adxectivo, grao non aplica, feminino, plural]; Tnfp [Relativo, non determinante, feminino, plural]; Vps30p [Verbo, presente, subxuntivo, 3ª, xénero non aplica, plural]

- d) por moi pouca utilidade que: P [Preposición]; Wm [Adverbio, modificador]; Idfs [Indefinido, determinante, feminino, singular]; Scfs [Substantivo, común, feminino, singular]; Tnfs [Relativo, non determinante, feminino, singular]

Temos claro, polo tanto, que non son locucións

- *por* + adxectivo/frase adxectiva + *que*
- *por* + adverbio/frase adverbial + *que*
- *por* + *máis* substantivo/frase substantiva + *que*
- *por* + *moito/a(s)* substantivo/frase substantiva + *que*
- *por* + *pouco/a(s)* substantivo/frase substantiva + *que*

Agora ben, que ocorre coas construcións nas que só aparece o cuantificador? Cremos que un primeiro paso necesario para avanzar no seu estudo é separar aqueles casos nos que o cuantificador amosa variación xenérica e numérica daqueles outros nos que aparece inmovilizado en masculino singular.

Así, sempre que *pouco* e *moito* se flexionan compórtanse coma se aparecese a frase nominal e o cuantificador desempeña unha función sintáctica na cláusula de relativo a través do pronome *que*. Nos exemplos seguintes, apréciase claramente a concordancia entre o cuantificador *pouca*, que a través do pronome relativo se converte en Atributo, e o Suxeito de *sexá* (*a diferenza entre as estacións*), e mais entre o indefinido nuclear *moitos* que, transferido mediante o relativo á función de Suxeito, concorda en número co Predicado da cláusula de relativo.

Se esta cativa variación térmica inflúe no metabolismo deses seres, terei que pensar que algúns deben incluso de hibernar, por pouca que *sexá* aquí tamén a diferenza entre as estacións. [CORGA: X. Bernárdez Vilar, *Big-Bang*, Xerais, 1996]

Nós seremos cincocentos de armas, e eles, por moitos que veñan, da dobra non pasarán. [CORGA: D. X. Cabana, *Cerco de ferro*, Nigra, 1994]

Nestes casos, naturalmente, seguimos sen estar ante locucións conxuntivas.

Non obstante, cando non hai flexión, os problemas para delimitar se o cuantificador *máis*, *menos* ou *pouco* funciona como antecedente do pronome relativo na cláusula que o integra difiren dependendo do elemento concreto do que se trate.

Así, nas ocorrencias con *pouco* dá a impresión de que se segue mantendo o valor cuantificativo, pois en todos os casos da versión 1.5 do CORGA os verbos que o acompañan acéptano e mesmo o requiren. Vexamos algúns exemplos representativos:

Hei constatar tamén, por pouco que me agrade, que os que nacen baixo o referido signo posúen unha cara grande e fea, pencosa e pustulenta, pouco pelo e este teixo ou vermello, un mirar aguzado e arrepiante, longo o pescozo, un nariz prominente e aberto nas súas ventas, a dentame brancuza e mal formada, escasa barba e o corpo algo chepado. [CORGA: X. Rábade Paredes, *A palabra secreta*, Galaxia, 1989]

Algún día, por pouco que furasen, as toupas chegarían alí, cada unha pola súa banda. [CORGA: X. Neira Vilas, *Papeis*, Transportes Castromil, 1992]

Por pouco que se queira a unha persoa... cómpre ser coma todos. [CORGA: P. Carballude, *¡Maldito D.N.I.!*, Xerais, 1998]

Nós, por pouco que che guste, tamén vivimos na casa do Minotauro. [CORGA: R. Loureiro, *O corazón portugués*, Galaxia, 2000]

E, en vez de maquetas de guerra, que as mesas soporten cuncas ateigadas de froitas exóticas, e ponme unhas cortinas, e encera ese chan, e deixa aí como tirados uns libros, e esquécete de mobilizacións e faille a cama con sabas limpas, que por pouco que Christina se ocupase da súa casa, polo menos en opinión de Eija-Liisa, non ía ser tan burra de non ter preparado o escenario que mellor lle axeitase para a sedución. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

A interpretación dos exemplos anteriores esixe a presenza do adverbio: *anque lle agrade pouco*, *malia que furasen pouco*, *aínda que se lle queira pouco a unha persoa*, *inda que che guste pouco...* Polo tanto, tampouco *por + pouco + que* será para nós locución conxuntiva.

Aínda que para delimitar e xustificar adecuadamente as diferenzas sería preciso realizar un estudo moito máis profundo do que podemos permitirnos aquí, cando o cuantificador é *moito* ou *máis*, parece que o seu grao de lexicalización varía e, en consecuencia, tamén o seu un comportamento. Sen saber moi ben en función de que se produce esa diferenza de gramaticalización, apreciamos ocorrencias nas que o

cuantificador funciona como tal na cláusula de relativo, exemplos de a)²¹³, onde non sería locución, e ocorrencias nas que se perde ou dilúe perceptiblemente todo significado cuantificador e só se percibe o concesivo, sendo imposible descompoñer, pois non se pode atribuír ningunha función con respecto ao verbo da cláusula que introduce, exemplos de b), onde si será tratado como locución conxuntiva:

1: Moito

a) valor cuantificador

O Pito Cairo: Non me é posibre, por moito que me esforce, comprender a vosa pasibilidade. [CORGA: E. R. Ruibal, *O cabodano. Cousas da morte. A sombra do bon cabaleiro*, Pico Sacro, 1977]

Desde aquela, por moito que o intentaba, en vez de ouro saíalle unha pasta verdosa e podre, de cheiro forte, que dañaba os ollos. [CORGA: C. Casares, *Os escuros soños de Clio*, Edicións do Cerne, 1979]

Eu arredei canto puiden e aturei sen berrar, pero non por estoico senón por medoñas, que por moito que lle temese á brasa do cigarro máis medo tiña a que me descubrisen. [CORGA: M. Riveiro Loureiro, *Premios Pedrón de Ouro 1985/1986*, Edicións do Castro, 1987]

A maior fíxo Farmacia e a máis nova inclinouse por Maxisterio, por moito que seu pai, e algo a súa nai, teimasesen orientala á medicina. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]]

Levanta e baixa os brazos, estorcíllate e destorcíllate sobre os calcañares, pois por moito que fagas e que desfagas, por moito que badues e deixes de baduar, por moito que acenes e desacenes, as de aguantar e de foderte igual. [CORGA: X. Fernández Ferreiro, *A ceo aberto*, Edicións do Castro, 1981]

E, no entanto, non sei por que se me antolla que, por moito que reza, fáltalle espiritualidade. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

b) valor non cuantificador

O xogo e, polo mesmo, un modelo da realidade social en que se produce, por moito que un xogo seña importado, por moita que seña a dependencia dun pobo neste senso, si a cultura e forte noutros ordes, éstes aparecerán sempre como peculiaridades do xogo, como un xenio especial que se manifesta con pulos. [CORGA: B. Cores Trasmonte, *Prólogo a Xogos infantiles de Galicia*, Follas Novas, 1979]

²¹³ Entre os de *moito* destaca pola súa claridade cuantificadora, marcando o contraste *moito* fronte a *algo*, o exemplo de *De corpo enteiro*.

Non, mamá, non, ti non estabas libre de pecado, nin eras trigo limpo, ti participabas, eras cómplice, gustabas daquelas festas que non podo imaxinar, gustabas por moito que ó día seguinte foses exculpar o pecado ó confesionario. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

E outras mulleres namoradas faranlle o traballo de cada día a estes figuróns, e as miñas simpatías, por moito que eles sexan sabios e boas cabezas, son máis delas ca deles. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

2: Máis

a) valor cuantificador

Doutro xeito mataranvos a todos, por máis que sexades, e arrasarán o voso quimbo. [CORGA: X. M. Martínez Oca, *As florestas do Mañuema*, Xerais, 1988]

Muller: Con esa xente que por máis que teimes en negalo sabe que tes pauto co demo. [CORGA: R. Vidal Bolaño, *Bailadela da morte ditosa (Sete baileretadas de amor e unha de morte)*, Sotelo Blanco, 1992]

Cronista: E, por máis que preguntaron, ninguén sabía darlles razón de onde nacera o Rei que buscaban. [CORGA: X. M. Carballo Ferreiro, *Teatro para o Nadal*, Citania de Publicacións, 1995]

Esther: Por máis que calcules as cousas ata no máis mínimo detalle, sempre hai algo que sae mal. [CORGA: R. Vidal Bolaño, *Rastros. Desgracia en catro tempos e unha presada de recordos*, Edicións Positivas, 1998]

Amedio: por máis que limpes sempre queda algo. [CORGA: C. Pazó, *Binomio de Newton / García*, Baía, 2006]

Ernesto: por máis que pases a fregona sempre queda algún. [CORGA: C. Pazó, *Binomio de Newton / García*, Baía, 2006]

b) valor non cuantificador

Carme: ¡Por máis que teñades pauto co demo non alonxaredes a xusticia desta terra! [CORGA: E. R. Ruibal, *O cabodano. Cousas da morte. A sombra do bon cabaleiro*, Pico Sacro, 1977]

Segue en voz baixa, por máis que ben oíble para todos. [CORGA: Ú. Heinze, Anaiansi, Ir Indo, 1989]

pero agora xa non son unha forza, por máis que poidan facer aínda algunha que outra barbaridade. [CORGAetq: ANT2008-07-10/84]

E eu toda a noite preocupado polo momento en que me vise obrigado a recoñecer que me chamo Ernesto (moi visto iso da importancia de chamarse tal, pero ninguén resiste a tentación

por máis que a maioría nin puta idea de que en inglés o xogo de palabras é a importancia de chamarse importante). [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Sentiuse confortado por iso, por máis que fose unha parvada, sentir que os cans da casa dos avós o recoñecían facíao sentir ben. [CORGA: M. Regueira, *L'affiche rouge*, Xerais, 2007]

Non é de extrañar, desde logo que non, explíqueille cun sorriso, e convideina a probar da nosa caña, feita na casa, por máis que tamén lle podía facer café se o prefería. [CORGAetq: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Nesta última secuencia, por exemplo, o valor cuantificador desaparece totalmente e só destaca o concesivo. Nela non inferimos que xa fixeran café e se esgotou, motivo polo que existe a posibilidade de facer máis cantidade, senón que se non lle gusta a augardente, tamén lle poderían facer café.

A gramaticalización desta construción en locución conxuntiva concesiva é total en casos como:

O pobre, ademais deste rol de intermediario co Máis Alá, é tamén, por máis que sui generis, un membro da comunidade. [CORGA: M. Gondar Portasany, *Crítica da razón galega. Entre o nós-mesmos e o nós-outros*, A Nosa Terra, 1993]

En ningún caso ese perigo (por máis que certo) de extinción pode esgrimirse contra os intentos de comunicación internacional. [CORGA: T. Moure, *Outro idioma é posible. Na procura dunha lingua para a humanidade*, Galaxia, 2005]

O remedio, por máis que viril e socorrido, non debeu dar resultado pois ao día seguinte tivo que gardar cama. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

En conclusión, a falta dun estudo moito máis rigoroso ca o que nós realizamos e coa información que nos achega a casuística das distintas ocorrencias de *por máis que* e *por moito que*, estas construcións serán tratadas en XIADA de xeito dispar:

- cando non se perciba valor cuantificador e non sexa posible atribuírle función, tratarémolas como locucións conxuntivas.
- cando se perciba valor cuantificador e sexa posible descompoñelas, atribuíndolle unha función con respecto ao verbo da cláusula na que se insire, serán desagregados os seus compoñentes en preposición, adverbio e relativo.
- en caso de as diferenzas non se apreciaren con claridade, optaremos pola caracterización como locución.

4.4.14.3. Locucións adverbiais

4.4.14.3.1. Pode que, poida que, pode ser que, poida ser que

Orixinada na perífrase modal de probabilidade co infinitivo *ser* (*poder + ser*), a expresión reducida *pode que/poida que*, inmovilizada na terceira persoa de singular do presente, ben indicativo ben subxuntivo, convive coa variante completa *pode ser que/poida ser que*, ambas as expresións sempre con verbo en subxuntivo. Así o demostra a lingua:

a) variante reducida

Pode que perda a batalla nesa maneira de interpretar competitivamente o mundo, pero vouno intentar. [CORGAetq: *GI1996-06/30*]

Están de moda, non sei por que, pode que sexa porque a nosa sociedade está máis enferma do que algún pensa pero... máis alá da metáfora fácil, o único certo é que as tramas sobre medicina arrasan na televisión. [CORGAetq: *GH2007-03-17/89*]

Estas razóns explicarían tamén a evolución do investimento das familias en vivenda, "cun ciclo expansivo que poida que alcanzase xa o seu punto culminante". [CORGAetq: *CG2000-11-07/80*]

a) variante completa

O que pasa é que en canto saia ao mercado pode ser que o prezo baixe". [CORGAetq: *ANT2005-03-03/57*]

Ricardo: Aí poida ser que as miñas verbas lle fagan mosa, que el parvo non é e eu seille dar por onde lle doe. [CORGA, M. Riveiro Loureiro, *O coronel ten a quen lle escribir*, Xerais, 1989]

Só se o verbo *ser* está explícito, malia que sempre en terceira singular, é posible conxugar o verbo *poder* nos distintos tempos, tanto de indicativo como de subxuntivo (excepto o futuro) e segue mantendo a significación de “é posible que”:

Podía ser que custase "un poquiño máis de traballo" pillalo, pero non había fallar: [CORGA, X. Alcalá, *O Larvisión e outros relatos*, Xerais, 1984]

Podería ser que o meu fose un esquecemento afectado. [CORGA, X. Alvilares, *Memorial de agravios*, Galaxia, 1992]

Medea: Poderá ser que ti xa non me coñezas, pero de seguro que os meus inimigos non me teñen olvidado. [CORGA: E. Alonso e M. Guede, *Medea. Versión libre sobre textos de Eurípides, Séneca e Anouilh*, Xerais, 1988]

Puido ser que andase corrida a voz de que tivera na man a vinganza do Gonzalo Novás e o deixara marchar sen facerlle dano. [CORGA: H. Puentes, *O bandido Casanova*, Xerais, 2000]

Frei Paspallás: Puidese ser que a música da gaita lle puxese a proa na testa e volvese de contado á terra nai. [CORGA: X. Lama, *Primeiro Concurso de Obras Teatrais Inéditas "Camiño de Santiago"*. Apartado A: *Obra de teatro profesional*, Hércules de Ediciones, 1994]

Ante esta situación temos dúas posibilidades de actuación: unha, considerar que *pode que/poida que* está gramaticalizada e se comporta coma unha locución adverbial; dúas, entender que son dúas unidades distintas (3ª persoa singular do presente de indicativo ou subxuntivo de *poder* e mais a conxunción subordinante *que*) nas que o significado de posibilidade achegado por *poder* na construción con subordinada substantiva de suxeito en subxuntivo permitiu elidir o infinitivo. Decidimos optar pola segunda das alternativas, a pesar de ser a primeira unha opción xa recollida nas gramáticas galegas (Álvarez e Xove, 2002: 350); RAG/ILG, 2003: 178 ou Hermida, 2004: 160), dado que:

1) A súa cohesión interna non é grande, pois pódense coordinar dúas cláusulas encabezadas por *que*:

Ás veces mesmo pode que os dous estean namoriscados e que os dous o ignoren; [CORGA: A. Conde, *Sempre me matan*, Ir Indo, 1995]

Poida que este non fose o seu mundo e que estivesen condenados a sentírense aquí eternamente extranxeiros, pero para o que lles compría sabían moverse e argallar ben. [CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

Poida que aínda vivan, que aínda garden no íntimo un lume ruín de rebeldía, pero, anque se agachen onde se agachen, aínda que mesmo teña agora que ir buscalos ó mais profundo dos infernos, hei velos mortos, desfeitos, esfandangados de vez e para sempre. [CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

No fondo poida que nos atopemos no bo camiño e que pronto se empecen a dar unhas condicións nas que a revolución chegue a prender. [CORGA: X. M. Martínez Oca, *Beiramar*, Xerais, 1983]

2) É posible intercalar elementos no seu interior, a semellanza doutras perífrases:

Pode tamén que a porta do departamento non peche. [CORGA: C. Gonsar, *Arredor do non*, Edicións do Cumio, 1995]

_E a política dos xornais, ese campo que tan ben coñeces e utilizas, tamén está entrando nunha liorta de inteseses brután, porque o voso, miña querida e sempre lembrada ex-muller, pode ser

engolido por outros máis grandes ou pode tamén que agora mesmo choque cos favores debidos a teu pai. [CORGA: L. Diéguez, *A canción do vagamundo*, Sotelo Blanco, 1986]

seguramente a informática sexa unha das profesións que máis esforzo de formación e actualización esixe, pode incluso que por enriba doutras como a medicina e a enxeñería. [CORGA: E. García Roselló, *Guía de Introducción á Informática*, Universidade de Vigo, 2000]

Temos por tanto que buscar unha transición máis rápida, capaz de xerar unha economía competitiva, e hai que empezar por facer que diminúa progresivamente o subsidio. [CORGAetq: *GII1996-06/30*]

Debemos tamén facer constar que nas zonas de contacto destas dúas entidades raciais existe unha mestizaxe grande. [CORGA: I. Caamaño, *Unha alternativa gandeira pra os montes da Galiza*, Galaxia, 1981]

O procesado, Enrique Escudero Dual, alias Conguito, que é reincidente, terá ademais que indemnizar con 119.000 pesetas a vítima, Celso Franco López, que xa está en liberdade. [CORGA: *CG1996-02-03/67*]

3) Coexiste coa variante inmovilizada en terceira singular de *poder* conxugado nos distintos tempos + *ser que*;

4) Non admite modificadores e semella comportarse coma *parece que/parece ser que*, combinación que ninguén se atreveu ata agora a clasificar como locución adverbial.

Na nosa opinión, é unha estrutura que a lingua habilita para expresar a posibilidade ou probabilidade, mais non constitúe locución adverbial, pois inda que se poida conmutar por *quizais* ou *talvez*, non está totalmente lexicalizada (conviven *pode que ~ poida que* ao lado de *pode ~ poida ~ podía ~ podería... + ser que*), non admite modificadores e semella comportarse coma *parece que/parece ser que*. É probable que esta decisión sexa cuestionable, mais os criterios empregados para a delimitación de locución (cohesión, significado e comportamento sintáctico) non nos permiten tomar outra decisión que non sexa a da súa análise como elementos desagregados.

4.4.14.3.2. Non obstante, sen embargo, con todo, así e todo, así a todo, en troques

Dado que a efectos de categorización é indiferente se as expresións que imos estudar nesta epígrafe encaixan dentro das estruturas concesivas ou das adversativas,

imos prescindir das súas similitudes e diferenzas. Interésanos, iso si, que en ambas se opoñen dous argumentos que levan a conclusións encontradas, e que nas adversativas o argumento que se introduce é moito máis forte ca nas concesivas, onde se presenta un obstáculo máis débil que pode ser superado.

Malia que na tradición hispana xa é un feito consumado o tratamento adverbial das locucións que encabezan este apartado, aínda non o é tanto na galega onde só comeza a aparecer. Así, dos manuais máis empregados, e a falta de que a Real Academia Galega publique a súa gramática, só Hermida (2004: 172) os considera adverbios conectores. Tanto as *Normas* (RAG/ILG, 2003: 183-184) coma Freixeiro (2000: 583) inclúenos na clase das conxuncións, mentres que Álvarez e Xove (2002) non acaban de concretar a categoría gramatical:

Teñen igualmente sentido adversativo as conxuncións copulativas (*e, e mais*) reforzadas por *sen embargo* e formas equivalentes (*nembargante(s), non obstante, así a todo ~ así e todo, con todo, por iso ~ por eso, emporiso ~ emporeso...*) [Álvarez e Xove, 2002: 655]

Non obstante, en relación coa relevancia que están adquirindo no terreo da lingüística actual os marcadores discursivos e o estudo da lingua dende unha perspectiva supraoracional, Freixeiro dedicou esforzos considerables a profundar na gramática do texto (2003) e na expresión da contraargumentación (2005a e 2005b) no galego escrito e, aínda que marxinalmente, si repara na categorización dos distintos conectores. Así, case coas mesmas palabras e coa mesma extensión, tanto no seu estudo parcial sobre algúns conectores contraargumentativos (2005a: 252) coma no monográfico sobre os marcadores discursivos e os conectores contraargumentativos en galego (2005b: 201-202), sinala que os conectores parentéticos como *non obstante, con todo, así e todo, no entanto*, etc., pertencen á categoría adverbial. É verdade que Freixeiro reproduce neses dous traballos as regras nas que se basea Portolés para realizar a distinción categorial entre conxunción e adverbio no marco dos conectores discursivos, mais é un avance considerable por canto implica para o galego a aceptación dos criterios diferenciais e a nova adscrición categorial dos elementos en cuestión.

No sistema de XIADA *non obstante, sen embargo, así e todo, así a todo, con todo* e *en troques* van ser tratadas como locucións adverbiais, por canto o seu comportamento corresponde ao dun adverbio e non ao dunha conxunción, xa que:

1) Son compatibles coas conxuncións:

García sinalou que o PP é "un rastrexador de popularidade", porque di o que quere oír en cada momento o auditorio ó que se dirixe e agora, "se hai que falar de auditorías externas pois fala de auditorías externas", pero sen embargo no seu programa electoral non consta. [CORGAetq: CG1996-02-14/89]

Acebes, que participou na Conferencia de Economía e Política Social do PP de Málaga, indicou que a Lei de igualdade efectiva de mulleres e homes, que foi aprobada con 192 votos a favor e a abstención do PP, ten " cousas positivas", pero, non obstante, conta con algo que "é claramente prexudicial para as mulleres e para a igualdade", que é "a sociedade cota", que é "como entende Zapatero a sociedade española". [CORGAetq: GH2007-03-17/35]

A vida era durísima, que aínda ás veces penso agora que xa non valía a pena de ser vivida, mais, con todo, íase levando adiante mentres houberse un algo de alento, cousa que a min non me faltaba a pesares de me atopar tan illado, moi lonxe dos meus compañeiros españois. [CORGA: M. Riveiro Loureiro, *Premios Pedrón de Ouro 1987*, Edicións do Castro, 1988]

Eis unha concepción da vida próxima ao estigma, que non acena buscando a compaixón do lector mais en troques consegue a súa solidariedade. [CORGA: ANT2005-02-03/48]

Sidorio, o acompañante de anos, guiaba nel e puña o contrapunto da súa voz raxada, contando os sucesos de toda Vilaquinte e mais contornos dunha maneira moi directa e clara, con ben menos requinte que o Acordeón, anque, así e todo, tamén con menos carne. [CORGA: H. Villar Janeiro e X. Rábade Paredes, *Morrer en Vilaquinte*, USC, 1980]

Pero, así a todo, sempre hai novidades. [CORGA: CG1996-02-04/8]

2) Poden aparecer en incisos:

A enquisa realizada entre consumidores e industriais destaca, non obstante, que por primeira vez desde agosto de 1995 a confianza xeral dos europeos na situación económica permanece estable. [CORGAetq: CG1996-02-08/94]

Nos países europeos da área da OCDE, sen embargo, a media situarase nun 49,1 por cento para 1996. [CORGAetq: CG1996-02-09/96]

"ERC, con todo, diríxese a colectivos moi concretos e CiU carece dunha política de aproximación aos inmigrantes", relata Àlex Milian. [CORGAetq: ANT2008-07-10/6]

En troques, vese incapaz de articular medidas de contención para afrontar a situación actual. [CORGAetq: ANT2008-07-10/14]

Así e todo, _rematou o dono da Pensión_ prometéu dar outra volta por eiquí; [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Así a todo, hai trousos nesta e noutras costas que non dan un peixe durante o ano; [CORGA: X. Rodríguez Vergara, *Redes e peixes. Saberes dun mariñeiro*, Xunta de Galicia, 2008]

3) Amosan a mobilidade correspondente a un adverbio, incluíndo nalgúns casos a posición final de enunciado, inda que esta, por incluírense no grupo semántico da contraargumentación, non é moi habitual. En todo caso, ocupan posicións imposibles para unha conxunción:

Rara vez é Juan un efe de frívolo, sen embargo; todo o máis é o efe das noticias. [CORGA: CG1995-02-10/53]

E dino, se cadra con bastante mais razón non obstante, os armadores coruñeses. [CORGA: TE1977-35/1]

Colombia, con todo, é outra cousa. [CORGAetq: ANT2008-07-10/82]

Pero Andrés non volveu mirar por ela, e ela en troques non facía máis que procuralo e agardar. [CORGA: F. A. Vidal, *Os pecados capitais*, Xerais, 2007]

Bo, xa eu non lle traía moita fê, pero así e todo... [CORGA: X. M. Martínez Oca, *A Negra / A caza das cascudas / As augas toldas / Terra sequeira*, Edicións do Castro, 1979]

Eu, así a todo _atalla o príncipe_, non son moi partidario de usalo. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

4.4.14.3.3. Aínda así

Non imos ocuparnos aquí da combinación casual do adverbio temporal (*aínda*) e o modal (*así*) cando se percibe con claridade neles a significación que lles é propia por separado. Por exemplo:

Sentouse no mesmo lugar contra o fondo, deixou o prato case sen probalo, non lle apetecía, tomou o café, un cigarro, logo outro. Ollou o reloxo. Outro cigarro. Unha camareira recollía as bandexas cos pratos. Xa está ben. Impaciente, cabreado consigo mesmo. ¿Vou para os corenta e aínda así? Xa está ben, por hoxe non fago máis o parvo, así e todo corenta anos case, a metade da vida, a partir de aí o declive, bo, non me amoles, nunca tan ben me vin. [CORGA: X. M. Martínez Oca, *Náufragos en terra. Os amores imposibles*, Xerais, 1995]

Recoñezo que son dos que, de ter que escoller, preferirían ter que se enfrontar á existencia dun home lobo, antes que á da convivencia no seo dunha sociedade tan indefensa, tan carente de sentido crítico e de formación cívica e humana, como a de mediados do s. XIX, susceptible de ser manipulada ata o extremo que moito me temo que o fora no caso de Manuel Blanco Romasanta. / ¿Viviremos no seo dunha sociedade aínda así? Pois non llelo sei. A opinión

reinante respecto do caso comentado é tan sólida que, a cantos lles preguntei, afirmaron que MBR foi un pobre home, inxustamente condenado. [CORGA: CG2000-12-21/81]

En ambos os casos a contigüidade non é forzosa e mantéñense as acepcións primarias, tal e como sucede con outros adverbios temporais ou mesmo noutra orde:

¿Pero aínda estás así? [CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Xa así as cousas, o morto incorporábase á comitiva de andantes para enceta-la peregrinación. [CORGA: F. Fernández del Riego, *San Andrés de Teixido*, Transportes Castromil, 1993]

É o costume de andar sempre así. [CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

As dúbidas que nos provoca esta combinación redúcense a unha soa: cando denota que a pesar das circunstancias expresadas se cumpre o enunciado na secuencia que introduce, debemos tratala como locución adverbial ou, pola contra, son dúas unidades distintas?

En favor da consideración como locución adverbial está a necesidade da presenza dos dous elementos para que o interlocutor infira que os feitos mencionados anteriormente conforman un argumento forte para conducir a unha conclusión contraria á expresada. Ex.:

- 1) Aínda así existe un gran desinterese pola música clásica en xeral, sobre todo, no público xuvenil... [CORGAetq: ANT2005-03-03/49]
- 2) Vilas e cidades dependen diso que se chama PXOM, pero aínda así estamos vivindo un crecemento desaforado dos nosos parques de vivenda. [CORGAetq: GH2007-03-17/17]

Suma así mesmo nesta opción o feito de que concorra maioritariamente a comezo de enunciado seguido polo sinal de puntuación “,” ou no interior da secuencia entre pausas, cun comportamento semellante ao doutras locucións adverbiais como poden ser *con todo* ou *así e todo*, que funcionan como marcadores discursivos:

- 3) Aínda así, máis do 80 por cento dos galegos carecen de luz eléctrica. [CORGAetq: ANT2008-07-10/57]
- 4) Nunca foi tan barato alugar un piso nesta cidade e, aínda así, fontes inmobiliarias sitúan en máis de 6.000 os que teñen dificultades para atopar inquilinos. [CORGAetq: ANT2005-03-03/53]

Xoga tamén no seu favor que, colocado despois dun signo forte de puntuación e seguido de pausa moi marcada, como conector discursivo de tipo contrargumentativo, non recibe modificadores nin pode ser negado, manténdose sempre inalterable. Así, das 15 ocorrencias que se dan na versión 2.4 do CORGAetq, 8 están en inicio de enunciado seguidas polo signo ortográfico “,” e 3 atópanse situadas en interior de enunciado, mais tamén entre comas; dos 4 casos restantes, 2, exemplos 5 e 6, aparecen en inicio e 2, exemplos 2 e 7, en interior sen pausa ortográfica marcada:

- 5) Aínda así existe un gran desinterese pola música clásica en xeral, sobre todo, no público xuvenil... [CORGAetq: ANT2005-03-03/49]
- 6) Aínda así amósase satisfeita co logrado desde unha secretaria que, co goberno bipartito, gañou peso político coa súa vinculación á presidencia da Xunta [CORGAetq: ANT2008-07-10/15]
- 7) Critícase o comportamento ruso en Chechenia, aínda así coa boca pequena, pero con Washington, cando chega a hora da verdade, todo é render unha preitesía que, mesmo, roza a humillación. [CORGAetq: ANT2005-03-03/65]

Na nosa opinión non reviste especial importancia a falta da pausa ortográfica nos dous primeiros exemplos deste último paquete, pois, malia que o máis común é empregalos seguidos de pausa, a independencia entoativa da que gozan non sempre se reflicte na escrita e non resulta estraño que en ocasións o marcador non estea situado entre signos de puntuación. Seguiríamos, pois, falando de locución adverbial cuxa función é discursiva. Sobre o 7 volveremos máis adiante.

A favor do tratamento como locución adverbial está tamén, a semellanza dos conectores discursivos de tipo contraargumentativo que acabamos de ver na epígrafe anterior, a liberdade posicional da que goza a expresión, pois pode aparecer en posición final de enunciado (exemplos 8 e 9), separando un Suxeito do seu Predicado (secuencias 10, 11 e 12), ou mesmo separando o Predicado do Complemento Directo ou Atributo (ocorrencias 13 e 14):

- 8) Pololo: Recadeiro, díxenche que era recadeiro para que non te entusiasmases, e aínda así... [CORGA: S. de Toro, *Unha rosa é unha rosa (Unha comedia de medo)*, Xerais, 1997]
- 9) _Amaro, non me deixes, non te arredes de min _pregoulle, aínda así. [CORGA: D. Ruiz, *Dentro da illa*, Galaxia, 2005]

- 10) Quen, aínda así, se arriscaba á desgracia de encontralo, conmovíase non só ante o seu aspecto desamañado e triste, senón, sobre todo, pola súa rotunda argumentación: [CORGA: A. Rodríguez Baixeras, *O rei dos ameneiros*, Xerais, 1987]
- 11) Por esta razón, moitos deixaran a misa do crego, pero outros, se cadra polo medo "a las medidas oportunas", dubidaran da utilidade de escoitar ó Coengo, que aínda así falara dous domingos mais; [CORGA: M. Suárez Abel, *A gran novela / Noite de luar / O home pequeno / O tesouro de Petro Xesto / O fillo de serafín / Anti-memoria dun día / ¿Pra qué dilatar a sua anguria? / O pau*, Edicións do Castro, 1982]
- 12) E o que tampouco debe desaparecer é, aínda así, o famoso contrato, aínda que sexa interno, para que os dereitos e obrigas sexan da compañía, independentemente das persoas físicas que a integren. [CORGA: P. Feijoo Barreiro, *¡Viva o fu remol! Ópera en cinco actos sobre a música ó pé da rúa*, Xerais, 2005]
- 13) Non hai, aínda así, revelación para os implicados. [CORGA: E. Veiga, *Final de película*, Deputación Provincial da Coruña, 2006]
- 14) Os progresos da produtividade nos servizos son, aínda así, ben reducidos. [CORGA: J. M. Mella Márquez, *Revista Galega de Economía: Cambio estrutural e eficiencia productiva da economía galega*, Universidade de Santiago de Compostela, 2000]

En contra da consideración como locución adverbial está o feito de que o adverbio inclusivo *aínda* non precisa do adverbio *así* para expresar valor concesivo, pois basta con que siga un xerundio (exemplo 15) ou un complemento modal (exemplo 16):

- 15) O que si vexo é que, aínda non habendo en Galicia un partido nacionalista tipo CiU, hai institucións que teñen hoxendía un peso autonómico moi grande. [CORGAetq: *GII1996-06/30*]
- 16) Os cálculos poden ter un erro que multiplique por catro esta estimación, pero aínda nese caso rebaixaría moitísimo a idade que se tiña por certa para as derradeiras erupcións volcánicas marcianas. [CORGA: *CG2000-11-20/47*]

De todas formas, o valor concesivo do xerundio non se consegue só cando o precede *aínda*, senón tamén noutras circunstancias. Así, é frecuente nas construcións absolutas con *podendo* (exemplo 17), coas locucións *e todo* (18) e *sen embargo* (19), e tamén precedendo o xerundio *nin* ou *nin sequera* (20 e 21):

- 17) Unha terceira variante deste grupo configúranas aquelas persoas que podendo falar galego en público, porque é a lingua da súa familia, non o fan por razóns ideolóxicas e políticas. [CORGA: *GH2008-08-06/111*]

- 18) Chovendo e todo, paráronse no medio e medio da Quintana dos Mortos. [CORGA: M. Rivas, *O lapis do carpinteiro*, Xerais, 1998]
- 19) Barreiro fixo fíncapé nos avances experimentados no ensino a prol do galego, salientando sen embargo que "parece que entran nas aulas máis galegofalantes dos que saen", o que vén a significar que "o sistema escolar non cumpre os obxectivos lingüísticos estipulados", dixo. [CORGA: CG2002-05-18/33]
- 20) Nin voando lle daría tempo pra ter chegado, fora por onde fora. [CORGA: X. M. Martínez Oca, *Número de patente / Querido Brais / A figueira*, Edicións do Castro, 1978]
- 21) Nin sequera rebaixando un pouco o prezo. [CORGA: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

O mesmo valor que estamos tratando para *aínda así* e nas mesmas circunstancias –en incisos, con xerundio ou con complemento modal–, tamén o achegan os outros adverbios inclusivos (*mesmo* e *incluso*, sendo *ata* moi raro), inda que a súa frecuencia de aparición na versión 1.5 do CORGA sexa moito menor (exemplos 22, 23, 24 e 25):

- 22) Entrevistado: Ó mellor, hai anos puido darse algún caso, aunque nunca se pode xeneralizar, que nalgún lugar mesturan a bolsa amarela e a negra e a xente vira que ía ó mesmo camiión. Incluso así, ó mellor despois separábase, dependendo do sistema de recollida. [CORGA: CG2000-12-18/6]
- 23) Pero, mesmo así, notaba a desorde no corpo e o corazón batendo velozmente. [CORGA: M^a. X. Queizán, *A semellanza*, Sotelo Blanco, 1992]
- 24) E isto sabiamos todos que non llo perdoaba, que ata así o dixo el en certa ocasión a pouco de acontece-los feitos. [CORGA: X. Fernández Ferreiro, Agosto do 36, Xerais, 1993]
- 25) Hai xente que, mesmo sendo boas persoas, non entenden nada. [CORGA: D. Ameixeiras, *Tres segundos de memoria*, Xerais, 2007]

Finalmente, para rematar coa casuística dos diferentes enunciados en que pode aparecer a combinación *aínda así*, quedáanos destacar cando é obxecto de negación (casos 26 e 27), cando modifica a un adxectivo (28?) e a súa combinatoria con outras expresións de tipo concesivo (exemplos 29, 30 e 31):

- 26) Non cho discuto, mais nin aínda así quixo deitarse contigo. [CORGA: A. Risco, *Tres situacións elementais*, Galaxia, 1991]

- 27) Pero nen aínda así saberán con certeza do frío louco, frío de pegarse na alma, frío de humidade cravada que estaba a sofrer don Pedro nesta que foi historia súa e nosa. [CORGA: L. Villalta, *As chaves do tempo*, A Nosa Terra, 2001]
- 28) Nunha parte da mesa, a menos porca, aínda así noxenta, dous pocillos agardaban silenciosos. [CORGA: X. Manteiga, *Morrer na herba*, Edicións Positivas, 1995]
- 29) Aínda así, e a pesar de que dixo todo iso, os empregados da oficina pensaron que tururú e mellor non andar a escribir parvadas, non vaia ser que alguén remate despedido por un casual. [CORGA: F. Castro, *Spam*, Galaxia, 2006]
- 30) Con respecto á mortalidade, gracias ós diagnósticos máis precoces, o índice reduciuse nos últimos anos, anque aínda así oscila arredor do 10 por cento e elévase ó 18 por cento en seguimentos de 30 anos. [CORGA: *CG1996-02-23/59*]
- 31) Ama: Aínda así e todo, aínda que o momento che puxera coitelos nas entrañas, eu sentín un xúbilo inmenso de te ter reencontrado. [CORGA: E. Alonso e M. Guede, *Medea. Versión libre sobre textos de Eurípides, Séneca e Anouilh*, Xerais, 1988]

Chegados a este punto da nosa reflexión non temos aínda claro se estamos ante unha única unidade ou dúas. Sabemos que se usa preferentemente ao comezo de cambio de turno conversacional, a xeito de marcador discursivo contraargumentativo, mais tamén que o significado denotado se expresa así mesmo cos demais adverbios inclusivos e que se precisa un deles seguido dun xerundio ou un complemento ou un adverbio modal. Evidentemente, non estamos cuestionando se é un marcador discursivo ou non, pois a efectos de categorización gramatical non nos interesa a resposta. Non obstante, parece que o seu comportamento se aproxima ao doutras locucións do seu mesmo tipo, aínda cando cremos que tamén hai argumentos para chegar á conclusión contraria (negación, modificación?).

Nos poucos traballos especializados existentes sobre este tipo de partículas que tocan esta combinación concreta para o castelán (Santos, 2003; RAE, 2009), os autores decántanse pola consideración de locución adverbial.

En galego, non gozamos aínda de ningunha obra que, a modo de dicionario, compendie as distintas locucións e nos explique as súas características e usos. Temos que recorrer aos poucos traballos parciais que sobre estas cuestións existen²¹⁴,

²¹⁴ Tamén Ferreiro (2002) e Domínguez Portela (2006) contribuíron co seu traballo á área discursiva, pero centrándose nunha parcela moi concreta: en *pero* o primeiro e en *pois* a segunda.

converténdose Freixeiro no único que abordou en xeral o estudo pragmático (2003) e o dos conectores contraargumentativos en particular (2005a e 2005b).

No artigo sobre os conectores adverbiais (2005a)²¹⁵, Freixeiro propón como posibles e compatibles dúas vías de gramaticalización para *aínda así*: a partir de *aínda así e todo*, con perda da última parte da expresión e a partir de construcións concesivas con xerundio. Afirmase, ademais, na gramaticalización de *aínda así* como

conector contraargumentativo parentético especializado en remarcar a grande forza argumentativa dun primeiro segmento e, porén, de modo similar ao que acontece con *así e todo* ou *con todo*, introducir unha conclusión contraria á esperábel dese. Igual valor ten *mesmo así*, que se pode considerar como unha variante de *aínda así*. [Freixeiro, 2005a: 266-267]

Nós estamos de acordo en que posúe un significado conxunto e que está cohesionada, malia que o significado final non difire en exceso do que lle proporciona a suma dos elementos que a compoñen. Instigados polo probable funcionamento como marcador discursivo en numerosos casos e mais polas facilidades que para o seu tratamento computacional supón unha única etiquetaxe, decidímonos finalmente pola súa inserción no sistema de XIADA como locución adverbial, mais sendo conscientes das súas anomalías. Por suposto, o mesmo tratamento seralle outorgado aos demais adverbios inclusivos seguidos de *así*: *mesmo así*, *incluso así* e *ata así*.

4.4.14.3.4. Nada máis

As dúbidas en relación con esta combinación xurdiron cando a atopamos no corpus a etiquetar seguida de infinitivo, pois non parecía que estiveramos ante dúas unidades. Por exemplo:

Ao final, todo saíu moi ben", asegura Rochela, o deportivista, nada máis aterrar en Barajas. [CORGAetq: DLV2007-05-15/108]

"En Rusia non hai democracia", dixo o escritor, un gran crítico do desaparecido presidente ruso Borís Eltsin, nada máis chegar a Moscova, desatando unha forte polémica. [CORGAetq: GH2008-08-05/44]

O xuíz Ruz, que se atopaba de garda o sábado pasado, iniciou as dilixencias de oficio nada máis recibir o informe policial sobre o seu desenvolvemento e pediulle que se lle remitan novos informes policiais polo acto da capital donostiarra. [CORGAetq: GH2008-08-05/55]

²¹⁵ A única diferenza entre o estudo parcial (2005a) e o monográfico (2005b) referente ao conector contraargumentativo *aínda así* é o emprego do conector aditivo *por outra parte* ao comezo do apartado e o cambio na numeración dos exemplos no estudo xeral.

A consulta de rigor ao DRAG (1997, 2012²¹⁶) confirmanos as sospeitas, mais non nos aclara se é unha locución conxuntiva ou adverbial:

Nada máis. 1. Tan pronto como. *Nada máis velo, acercóuselle para darlle unha aperta. 2.* Máis nada. *Non levaba nada máis ca unha maleta.*

Iso si, introduce novas variables para a locución: a) expresión coa que se remata unha enumeración ou relato, e b) *nada máis* ou *máis nada*, ningunha cousa máis (*Non quero máis nada. Non levaba nada máis ca unha maleta*).

O dicionario de IR INDO non nos axuda, pois non recolle a combinación; o Xerais, pola súa banda, refinou a definición dende a edición do 2000 á do 2009, mais tampouco dá conta de todos os usos:

2000

N. máis, *loc. conx.*, tan pronto como (*nada máis vernos, marchou correndo*).

N. máis ou **máis n.** ningunha cousa máis (*non precisaba máis nada*).

2009

N. máis, *loc. conx.*, tan pronto como.

N. máis / máis n. *loc.s.*, ningunha cousa máis.

Álvarez, Regueira e Monteagudo (1986: 260) só recollen entre as combinacións de *máis* con outros indefinidos a dupla variante: *nada máis ~ máis nada*, mais non lle asignan un valor especial. Freixeiro (2000: 291) non identifica unha unidade *nada máis*, pero si aclara que *nada* con verbos intransitivos ou como modificador dun adxectivo debe considerarse adverbio de cantidade (*Non corre nada / Non é nada torpe*), en tanto como CD é un pronome cuantificador (*Non quero máis nada*). Álvarez e Xove (2002: 486-487) volven incidir na ambigüidade de *nada* (Non ve *nada* [N: 'ningunha cousa'; adv. 'en absoluto'].) e nas dúas posicións posibles para *nada* cando vai acompañado de *máis*.

É curioso, non obstante, que en relación con esta unidade ningún deles repare na variante popular reducida *namais*, que claramente actúa como un adverbio:

²¹⁶ Dispoñible en <http://www.realacademiagallega.org/diccionario#inicio.do> [consultado o 28/05/2012]. Por erro non se inclúe na versión web a variable que nós denominamos a) mais é a que corresponde, xa que non tería sentido repetir como sinónimo a entrada mesma: "**Máis nada. 1.** Ningunha cousa máis. *Non quero máis nada.* SIN. **nada máis. 2.** Expresión con que se termina unha enumeración ou un relato. *E máis nada, aquí acaba o conto.* SIN. **máis nada**".

Descontando as del, había catro ringleiras namais, que a muller e os vellos aínda non pisaran a aira aquel día. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Namais cocín unhas pataquiñas en branco cuns chícharos de Erundina e unha ceboliña e listo. [CORGA: X. Borrazás, *Na maleta*, Sotelo Blanco, 2000]

_Si, muller, si; ¿quieres comer algo, anque só sexa un bocado namais? [CORGA: X. Vázquez Pintor, *A memoria do boi*, Xerais, 2001]

¿Quen dixo que só houbo a guerra dos Trinta Anos ou que namais vai haber a de Iraq? [CORGA: ANT2002-09-19/51]

E, por outra banda, tamén se bota en falta que non se indique, por un lado a relación entre *namais* e os usos paralelos de *nada máis* e, por outro, a proximidade que parece existir entre eses casos e mais os adverbios exclusivos *só*, *soamente*, *unicamente*... Referímonos, como non, a exemplos coma os seguintes nos que sería posible a conmutación por un destes últimos adverbios e nos que ou ben non é factible a anteposición de *máis* sen que a secuencia sexa agramatical ou ben, se o segue sendo, precisase unha puntuación distinta que muda a interpretación e sitúa a *nada máis* como elemento de peche discursivo:

Corenta quilómetros de traxecto nada máis. [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Cánto me gustaría sabelo todo sobre el, non ter o remorso de ir de falso e pedante, reprocharlle a Anxo que nada máis vira en min un futbolista. [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Foi un tempiño nada máis pero foime longo dabondo, que ás veces aínda tiña eu medo de ir dar na esquizofrenia. [CORGA, M. Riveiro Loureiro, *O coronel ten a quen lle escribir*, Xerais, 1989]

Cunha moeda miña, que se é túa nada máis ten cara. [CORGA: S. de Toro, *Unha rosa é unha rosa (Unha comedia de medo)*, Xerais, 1997]

Dado que semellan formar unha unidade cohesionada na que ningún dos dous elementos é prescindible, onde *nada* non pode ser cuantificador e *máis* o seu modificador, e onde o significado varía do da suma dos dous constituíntes, no corpus de adestramento decidimos tratalo como locución adverbial.

Optamos así mesmo por idéntica análise, locución de tipo adverbial, cando, coma nos exemplos iniciais, segue un infinitivo, entendendo que esta introduce unha cláusula temporal:

Nada máis verllo seu pai, deulle tal ataque de ira que llo arramplou das mans e llo guindou ó fondo da congostra, onde xusto foi caer de pé. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Verónica, nada máis percibir o cheiro a carne, abriu devagar os olliños, palpebreou atordada e devorouna nun intre. [CORGAetq: U. Casal, *Vidas exemplares*, Xerais, 2006]

4.4.14.4. Locucións prepositivas

4.4.14.4.1. Dentro de

En xeral, os distintos dicionarios (DRAG, 1997 e 2012; Xerais, 2000 e 2009; Ir Indo, 2004), así como as *Normas* (RAG/ILG, 1995: 179; 2003: 180), consideran *dentro de* unha locución de tipo prepositivo, tanto no seu valor temporal coma no locativo.

Obviamente, unha solución como a adoptada polas obras lexicográficas sería a máis cómoda para nós e non daría lugar a erros na etiquetaxe automática. Non obstante, cremos que iso se afastaría das características que perseguimos a anotación que realizamos: rigor, exhaustividade e consistencia na etiquetación.

A pesar de que *dentro de* e *en* alternan en numerosos contextos, o valor da preposición é moito máis amplo e inespecífico ca o do adverbio. Así é posible conmutar case sempre *dentro de* por *en*, pero non sempre é correcto o contrario: *Estou dentro da [na] casa* pero *Estou deitada na [*dentro da] herba*.

Sen dúbida, *dentro* seguido da preposición *de* e dalgún substantivo ou expresión temporal forma unha locución prepositiva que expresa o tempo que falta para que algo teña lugar. Vexamos algún exemplo:

Nunhas xornadas sobre o Sector das Telecomunicacións organizadas polo Instituto de Estudos Superiores da Empresa (IESE), o directivo sinalou que dentro de cinco anos a facturación do sector aumentará ata os 16.000 millóns de euros, un 9% anual máis. [CORGAetq: CG2002-05-30/14]

O ministro de Enerxía de Venezuela, Alí Rodríguez Araque, anunciou que a Organización de Países Exportadores de Petróleo (OPEP) podería realizar un recorte na súa produción de

500.000 barrís diarios dentro de 10 días se o prezo medio do barril se coloca por debaixo dos 22 dólares. [CORGAetq: CG2000-12-24/44]

Dentro de pouco os Quince terán que levantarlle tódalas restriccións á competencia en materia de telefonía móbil que queden vixentes. [CORGAetq: CG1996-02-18/73]

Nestas secuencias obsérvase que o adverbio non sinala localización espacial senón temporal. O grao de cohesión da unidade complexa demostrase coa coordinación dos termos (exemplos 1 a 3), malia que tamén sexa posible atopar algunha ocorrencia, as menos, coa repetición da preposición *de* (exemplos 4 e 5), ou mesmo coa repetición de toda a locución (6):

- 1) Prognostican que dentro de cinco ou sete anos se incrementará significativamente o número de agresores sexuais de menores. [CORGAetq: CG2002-05-30/14]
- 2) Non te bañes ata dentro dunha hora ou dúas, ata que acabes de facer a dixestión. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]
- 3) Centro Nacional de Investigacións Oncolóxicas (CNIO), que estará plenamente operativo a mediados do 2001, ofrecerá os seus primeiros resultados "reais e tanxibles" para a sociedade dentro de cinco ou dez anos, dixo o seu director Mariano Barbacid. [CORGA: CG1999-05-30/47]
- 4) Flor Maceiras: Esas mesmas persoas dentro de tres meses ou de seis meses poden xa estar noutra situación e, o que é peor, non hai garantía de que siga habendo series de televisión nin de que siga habendo Centro Dramático, nin de que siga habendo nada como para garantir sequera algo. [CORGAetq: TDN1996-02/2]
- 5) Desa maneira esquecerías a ese Conde, que... volve dentro duns meses ou dun mes, nada se sabe aínda, completa Anaïansi a frase. [CORGA: Ú. Heinze, *Anaïansi*, Ir Indo, 1989]
- 6) A doenza no menisco externo da súa perna dereita está aí e provocará unha operación que pode ter lugar dentro dunha semana ou dentro de dez anos. [CORGA: CG1996-02-17/105]

No noso sistema os casos anteriores etiquétanse sempre como locucións de tipo prepositivo. Porén, cando a significación non é temporal, xorden as dúbidas. Cremos que para explicalo se podería recorrer á metáfora dunha liña que se inicia co valor claramente locativo e que vai progresando ata chegar ao de elemento relacionante ou introdutorio.

A posible paráfrase de *dentro de* por *en* non xustifica, no noso parecer, a uniformización da súa análise. Semanticamente, a diferenza que cremos subxace é que con *dentro de* seguido de substantivos que posúen unha parte exterior e outra interior se

especifica a interna sempre, mentres que con *en* podemos referirnos tanto á superficie, como á totalidade do obxecto referido, coma ao seu interior. Funcionalmente, varios son os argumentos que nos levan a non tratala como locución prepositiva cando o significado é de localización espacial. Vexámolos:

1) A segmentación [dentro] [da casa], que se corresponde coa etiquetaxe de *dentro* como adverbio seguido dunha preposición, mais artigo e substantivo –[dentro Wn dentro] [de P de] [a Ddfs o] [casa Scfs casa]– permite explicar a ausencia do complemento que encabeza a preposición *de*, recuperable contextualmente.

O contrario, a segmentación [dentro de] [a casa] –polo tanto locución prepositiva– implica a distinta categorización en función de se *dentro* aparece ou non modificado: adverbio, se non o está, e locución prepositiva se si o está. En principio, isto levaríanos á inconsistencia da etiquetaxe por non aplicarlles a outros elementos en circunstancias parecidas a mesma análise. Por exemplo, os pares seguintes:

O gato está dentro. / Estou lonxe. / Ven despois. = Wn

O gato está dentro da casa. = Lp0

Estou lonxe da casa. / Ven despois do xantar. = Wn + P...

2) É posible omitir un modificador, mais non o termo dunha preposición: *O gato está dentro pero *O gato está dentro de*.

3) *Dentro* englobase no reducido grupo de adverbios que admiten complementos introducidos por *de*, tales coma *fóra, diante, detrás, antes, despois, cerca, lonxe...* Todos eles denotan ubicación e admiten modificadores de natureza focal coma *xusto, xustamente, incluso, mesmo*, ou a interpolación dalgún elemento discursivo entre eles e o modificador con *de*:

Mais non fago outra cousa que vos repeti-lo que se comenta en Amesbury, dentro incluso dos muros da abadía. [CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]

Dentro xa do portal, ó pisar as escaleiras brillantes de cera que renxían baixo dos nosos pés, volteime ver neno, da man da miña nai, sobindo naquela penumbra e un medo tolo, un pánico irracional empezou a se apoderar de min. [CORGA: X. M. Martínez Oca, *Número de patente / Querido Brais / A figueira*, Edicións do Castro, 1978]

Primeiro, porque a revista ten altura e espesura cultural, dentro, por suposto, do segmento ¿dise así? de lectores potenciais, forzosamente reducido, xa que, editada en Cataluña por e

para galaicoparlantes, é claro que o ámbito de difusión ten que cingirse, necesariamente, ó dos centros e asociacións frecuentadas por macos, charnegos e demais galegos establecidos na autonomía catalana. [CORGA: *CG1995-02-23/2*]

Detrás mesmo do asento da mestra, colgado da parede, estaba un retrato moi grande dunha estraña muller que tiña unha balanza na man, un león ó lado e detrás a bandeira tricolor da República. [CORGA: X. Alvilares, *Memorial de agravios*, Galaxia, 1992]

4) A presenza do posesivo feminino singular, como equivalente de *de + pronome persoal*, máis propio dos substantivos ca das preposicións, modifica o adverbio indicando o punto de orientación:

Como vou medir ca miña escala de importancias a importancia das que viven dentro túa? [CORGA: X. M. Martínez Oca, *Beiramar*, Xerais, 1983]

Unha sorte de ensoño envolveu a Miguel, que sentiu dentro súa, por primeira vez en moitos meses, unhas tremendas ganas de respirar. [CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

5) Finalmente, a posibilidade de establecer unha gradación cos adverbios comparativos:

O seu brillante elenco de tiradores, reboteadores e demais, actuando como o facía a excelente altura, necesitaba toda a súa capacidade de concentración porque o TDK estaba tan dentro do partido como o mesmísimo favorito. [CORGA: *CG1996-02-26/123*]

Non só lle disgustaba que lle andasen a fozar na moa senón que a súa posición, tan dentro da boca, acentuaba o desagradable da situación. [CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

Era un home calado, coma se vivise noutro mundo, ou moi dentro de si mesmo; un home diferente de nós todos. [CORGA: X. Neira Vilas, *Galegos no Golfo de México*, Edicións do Castro, 1980]

Nacemos desta boira que se volveu fantasma, procesión de defuntos ou cadea de sangue que vén de moi atrás, e tamén de moi dentro e moi abaixo. [CORGA: X. Rábade Paredes, *Como levar un morto*, Sotelo Blanco, 2001]

Así pois, cando o seu referente é espacial, cando o adverbio *dentro* se opón a *fóra*, malia ir acompañado por unha frase preposicional con *de*, cremos que non se debe considerar locución prepositiva senón un adverbio, que neste caso está modificado por unha frase preposicional con *de* (exemplos 7 a 10), mais que tamén podería aparecer como único integrante da frase adverbial (exemplos 11 a 14), modificar el mesmo a un

deíctico (15 a 17), ou ser o núcleo da frase adverbial modificada á súa vez por unha prepositiva que está actuando como modificador dun deíctico locativo (18 a 20):

- 7) En poucos segundos, só o bruar; ningún ser vivo dentro do circundo de pedra. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]
- 8) Poder falar, ou mesmo ver, dentro do establecemento ao cociñeiro de Reses Bravas, Marcial, sen apelido coñecido, é practicamente imposible. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]
- 9) No pacto cos empresarios, o PT conseguiu inmiscuílos no desenvolvemento interno para que a oligarquía invista dentro do Brasil. [CORGAetq: *ANT2005-03-03/62*]
- 10) Poder falar, ou mesmo ver, dentro do establecemento ao cociñeiro de Reses Bravas, Marcial, sen apelido coñecido, é practicamente imposible. [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]
- 11) Unha vez dentro, sabes con quen podes contar e, ó mesmo tempo, ti tamén comezas xa a contar para os demais". [CORGAetq: *TNI997-04/7*]
- 12) Cando escribo, poño a lidar o optimista e o pesimista que levo dentro, a ver que pasa. [CORGAetq: *ANT2008-07-10/27*]
- 13) A última vez que sentín a súa mirada a remexerme as carnes por dentro foi aquela mañá. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]
- 14) Quería botar pestes, protestar, berrar, pero tódalas miñas emocións sentimentais quedaron atrapadas dentro. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]
- 15) Levanteime da cama pouco a pouco, e o enfermo aquel axudoume a baixar as escaleiras, senón eu finaría tamén alí dentro. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]
- 16) era o uniforme dos boqueras e a maneira de se pasear e de collerte polos pelos e de che estampar a cara contra a porta e de arrearche patadas e de te empuxar polas escaleiras e de cachearte e de mofarse de ti preguntando que será iso que cae ó chan de entre as sabas que levas na bolsa despois do vis a vis, e divírtense á túa conta e trátante coma un can porque te odian e porque saben que eles si están condenados a pasar toda a vida alí dentro porque non serven para outra cousa e cando por fin che dan a bóla non o soportan e cómeos o demo e teñen que soltarche iso de que tarde ou cedo xa volverás. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]
- 17) Si, certamente, aló dentro semellaba haber sangue callado. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]
- 18) Aínda foi ben que o garda do celeiro da pólvora non estaba quentándose na forxa e non houbo que chorar ningunha morte, pero no campamento rebentaron os vidros todos e desgonzáronse as portas

que saíron voando polo aire e tras delas todo canto había alí dentro dos chabolos. [CORGA: H. Puentes, *O bandido Casanova*, Xerais, 2000]

19) Logo, xa un pouco máis calmada, máis dona de min, vestínme e dispúxenme a saír, alí dentro daquel apartamento non podía aturar máis tempo. [CORGA: X. M. Martínez Oca, *Como paxaro caído do niño*, Baía, 2006]

20) Pareceulle que non brillaba tanto coma alá dentro da cova, pero cecais o reflexo da lua... [CORGA: X. M. Martínez Oca, *A gran novela / Noite de luar / O home pequeno / O tesouro de Petro Xesto / O fillo de serafín / Anti-memoria dun día / ¿Pra qué dilatar a sua anguria? / O pau*, Edicións do Castro, 1982]

O diferente comportamento de *dentro de* con valor espacial obsérvase tamén nas coordinacións. Así, cando se coordinan os adverbios *dentro* e *fóra*, a frase preposicional modifica a ambos:

Existen en Galicia iniciativas empresariais modélicas dentro e fóra do país. [CORGAetq: *GI1996-06/30*]

Tui é a cidade que acollerá dende o vindeiro 10 ao 13 de marzo a primeira edición do Play-Doc "sen o glamour do cine nin a sobriedade dos que escriben para eles mesmos", un festival pioneiro adicado exclusivamente a proxección de documentais de todo o mundo, que se quere converter en "referente obrigado no panorama audiovisual dentro e fóra do Estado". [CORGAetq: *ANT2005-03-03/73*]

María Rosario foi a secretaria persoal de Paulo; sempre solteira e fiel ao xefe dentro e fóra do horario de traballo, que para ela nunca tivo límites. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Un día na vida de Iván Denísovich valeulle a Solzhenitsin o recoñecemento dentro e fóra da URSS, pero o éxito non evitou que o KGB o continuase asexando. [CORGAetq: *GH2008-08-05/44*]

Ou tamén cando están en correlación, onde cada un deles pode posuír un modificador diferente:

A realidade é que tanto as dificultades que veñen de fóra, como as que se detectan dentro da casa, revelan a inexistencia dunha estratexia clara na que estivesen ben definidos os nosos intereses colectivos, as canles e a vontade política coa que darlles cobertura. [CORGAetq: *TN1997-05/1*]

Participo da firme convicción de que dentro da historia entenderemos cáseque todo e fóra dela incomprenderemos, tamén, cáseque todo. [CORGA: *ANT2007-02-08/25*]

Ata aquí, fóra os exemplos iniciais de locución prepositiva con significación temporal, é bastante claro que estamos ante un adverbio. Agora ben, cando non temos correlación nin coordinación con *fóra* e a referencia é ao interior non dun obxecto real senón dun figurado, a diferenciación entre a locución prepositiva e o adverbio xa non está tan clara. Na maioría dos casos non pode prescindirse do complemento con *de* nin *dentro* admite gradación, funcionando máis ben como elemento introdutorio ou relacionante, semellante a “no marco de”, “en relación con”, etc. Estes casos decidimos marcalos como locucións prepositivas no CORGAetq. Sirvan de mostra os seguintes exemplos:

Ademais, un veciño de Vigo que responde ás iniciais A.S.A. foi imputado onte por distribuír e posuír imaxes de pornografía infantil a través de internet, dentro dunha operación contra a corrupción de menores que se desenvolveu en oito provincias españolas e supuxo a detención de nove persoas e a imputación doutras once. [CORGAetq: GH2007-03-17/36]

Dentro das curtas, ademais de Café con leite, do irreverente Nicolás Guillén Landrián, poderá verse Untertage na que Jiska Rickels propón unha viaxe até as profundidades da terra, nas que desaparecen as cores, os sons distorsiónanse e as vidas dos mineiros póñense diariamente en perigo. [CORGAetq: ANT2005-03-03/73]

Dentro da demanda interna, o compoñente máis dinámico foi o investimento, que medrou en termos reais un 9 por cento, case o dobre que o inicialmente previsto, mentres que o consumo privado só aumentou un 1,8 por cento. [CORGAetq: CG1996-02-08/96]

A conferencia celebrouse dentro do VIII Simposio sobre a Economía catalana e a Unión Europea organizado por Banca Catalana e o Padroado Catalá Pro Europa. [CORGAetq: CG1996-02-14/114]

En caso de dúbida, e precisamente pola súa propia existencia, optamos por considerar que o termo non marcado é a locución prepositiva. A falta de estudos específicos sobre esta unidade en concreto e sobre o grupo no que se integra en xeral, fundamentámonos en dous criterios que cremos básicos para decidirmos polo adverbio: a validez ante a supresión da frase preposicional con *de*, e a posibilidade de modificalo cun adverbio de grao. Toda ocorrencia que non responda positivamente aos dous puntos anteriores será etiquetada no procesamento manual como locución preposicional.

4.4.14.4.2. Arredor de, ó redor de

En XIADA *arredor ~ ó redor* seguido da preposición *de* e dalgunha expresión numérica ou temporal conforma unha locución prepositiva, cuxo valor é aproximativo. Vexamos algún exemplo de cada:

Ademais, o delegado de Sanidade confirmou os enormes gastos sanitarios que causa a doenza, en total, un 2% do gasto sanitario anual, ó redor dos 2.000 euros ó ano por paciente co que o total se situou no 2004 nuns 60 millóns de euros. [CORGAetq: GH2007-03-30/31]

Con máis de 740 tendas, está presente en 21 países e emprega ó redor de 12.000 persoas. [CORGAetq: CG1999-05-29/14]

Tello sinalou que agora o Concello só dá cobertura a 400 persoas, cando o censo oficial indica que hai na Coruña ao redor de 8.000 con necesidades especiais que requiren estas axudas. [CORGAetq: DLV2007-05-15/25]

Cada ano acaban a carreira en Galicia arredor de cen xornalistas. [CORGAetq: ANT2005-03-03/28]

Segundo informou onte o Instituto Armado, a detención produciuse ó redor das 11.30 horas de onte despois de que os axentes do Destacamento Fiscal da Garda Civil e do Servizo Cinolóxico do Instituto Armado, apoiados no labor dun can detector de drogas, descubrisen en varios dobres fondos dunha maleta desta muller a substancia estupefaciente referida. [CORGAetq: GH2008-08-05/19]

Un accidente por alcance entre dous vehículos rexistrado na AP9 de Vigo colapsou preto dunha hora a ponte de Rande (Vigo), pero a circulación quedou totalmente restablecida arredor das tres da tarde. [CORGAetq: GH2008-08-05/25]

Non obstante, non a consideramos locución prepositiva senón adverbio –locución adverbial no caso de *ó/ao redor*–, seguido por unha preposición cando fai referencia ao espazo, real ou figurado, que rodea á persoa ou cousa que se expresa. Igual que sucedía con *dentro* no subapartado anterior, ou cos demais adverbios que se engloban no mesmo grupo, *arredor* pode aparecer só, modificado por unha frase preposicional con *de* ou acompañado por un posesivo non determinante. Representan esta análise os seguintes exemplos do CORGAetq:

Por moito que mirei arredor non había rastro da señora detrás da vidreira. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

"Alí, en Palencia, xa digo, obrigábanos a andar arredor e sen falar con ninguén, nin mirar para atrás sequera. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

para un xantar de festa xúntase a familia arredor da mesa do comedor. [CORGA: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Estas tres comarcas xiraron durante moitos anos arredor do sector do metal, na variante principal dedicada ao naval e a súa industria auxiliar. [CORGAetq: *ANT2008-07-10/70*]

No tocante á organización do poder galego, a proposta presenta varias novidades no que se refire sobre todo ao poder xudicial, concebido arredor dun Consello Xudicial de Galiza que represente o órgano de goberno dese poder en Galiza. [CORGAetq: *ANT2005-03-03/2*]

Xa se xuntara, ademais, arredor de nós, unha chea de curiosos, algúns deles con cara de posíbeis clientes, que un cos anos xa os vai coñecendo ó primeiro golpe de vista. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Cunha formación que axude se non a compartir, si a comprender os cambios que se producen arredor nosa, en lugares diferentes e á vez en tempo real. [CORGAetq: *GII1996-06/30*]

4.4.14.4.3. Ademais de ~ amais de, aparte de ~ á parte de

Igual que facemos con outros casos nos que un adverbio seguido dunha preposición é considerado locución preposicional na maioría dos manuais, en XIADA optamos pola análise desagregada, pois cremos que non conforman en realidade verdadeiras locucións, senón que a frase preposicional encabezada por *de* – neste caso en particular pero tamén en case todos os demais: *arriba de*, *debaixo de*, *despois de*, *detrás de*, *diante de*, etc.– só modifica, complementa o adverbio.

Hoxe en día este tratamento non é xa inusual. Mesmo é o único que Álvarez e Xove (2002) recomendan para os casos que tratamos:

As locucións prepositivas teñen a súa orixe en sintagmas nucleados, sobre todo, por substantivos e adverbios, que se foron fixando como perífrases para un uso nexual e, asemade, perdendo o seu carácter orixinario. Por tanto, só se debe falar de locución prepositiva nos casos en que o sintagma non sexa analizable como 'frase + prep'. [Álvarez e Xove, 2002: 631-632]

Polo tanto, nin no lexicón introducimos unha unidade de tipo preposicional para os segmentos desta epígrafe nin no corpus de adestramento se analizan como tales. A análise para *ademais de ~ amais de* e *aparte de* é sempre Wn (*Adverbio nuclear*) seguido de P (*Preposición*), en tanto que para *á parte de* é La0 (*Locución adverbial subtipo non aplica*) mais P (*Preposición*). Vexamos algún exemplo:

Este centro, ademais de achegarlle formación ás persoas con discapacidade, terá tamén unha área de formación para profesionais e os medios necesarios para a utilización das novas tecnoloxías. [CORGAetq: GH2007-03-17/32]

Por primeira vez, unha lei galega fixa unha deducción específica por fillo ademais das que establece o Estado. [CORGAetq: CG1999-05-01/22]

Schiffer diseña os seus propios calendarios, cos que gana entre 36 e 50 millóns de pesetas por ano, amais de protagonizar programas especiais na televisión alemana, que lle reportan entre 18 e 36 millóns de pesetas. [CORGAetq: CG1996-02-23/132]

Amáis da edición de libros, de Rompente salienta a vontade anovadora, dende a publicación das Follas de Resistencia Poética aos recitais poético-visuais (performances cando non se coñecía o anglicismo). [CORGAetq: GH2008-08-05/47]

Hai uns cantos que estamos convencidos de que, aparte do que fan os partidos _mobilizar os mentalizados_ hai un gran sector que hai que facelo galego a través da acción cultural, cunha política de pedagogía galeguizadora. [CORGAetq: ANT2008-07-10/22]

Aparte disto, novos informes económicos advirten das serias dificultades polas que atravesan as familias e determinados valores como a vivenda. [CORGAetq: ANT2002-10-10/36]

á fin e ao cabo subsisten importantes diferencias de salarios e prezos no propio mercado español que, á parte dunha moeda común, goza dun marco fiscal e laboral homoxéneo. [CORGAetq: TN1998-05/18]

A xestión do programa encoméndase ás respectivas sociedades de capital risco, NORPEDIP pola parte portuguesa e SODIGA por parte galega, que, á parte de contaren cunha dilatada experiencia en promoción industrial, veñen mantendo estreitas relacións de cooperación. [CORGAetq: GI1996-06/34]

Con todo, cremos que sería moi interesante poder afondar no comportamento do adverbio *ademais* e no da locución adverbial *á parte* cando van seguidos por frase preposicional con *de*, poñelos en relación coas partículas denominadas exceptivas e tamén cos adverbios inclusivos (*incluso, ata*) ou exclusivos (*nin, nin sequera*) cando aparecen na secuencia en relación cun complemento análogo ao que delimitan.

Por desgraza, nós aquí non podemos desviarnos do camiño trazado con este tipo de cuestións, mais si pensamos que pode ser bo deixar constancia daquelas unidades que, polas peculiaridades que as caracterizan, deberían ser obxecto dun estudo máis polo miúdo. É por isto que non queremos pechar este apartado sen dar conta dalgunha das características que percibimos ao observar os resultados de *ademais de*:

▪ Pode ir seguido de calquera preposición:

En terra, axentes das Forzas de Seguridade do Estado detiveron, ademais de a Carballa Magdanena e Antonio Mascato, a outras 7 persoas. [CORGA: CG1995-02-08/34]

Actualmente a Universidade de Santiago conta con 40.000 estudantes, distribuídos entre os campus de Compostela e de Lugo, ademais de con 1.800 profesores e preto dun milleiro de traballadores da administración e de servizos. [CORGA: CG1995-02-04/35]

Segundo o Instituto Xeográfico Nacional, o movemento sísmico sentiuse, ademais de en Becerreá, noutras localidades lucenses da zona de Sarria. [CORGA: CG1996-02-04/51]

• Posúe mobilidade: tanto pode antepoñerse ao elemento co que se relaciona, como pospoñerse, como aparecer separado del:

Ademais de si mesmo _o que fai con notable ecuanimidade_ Voltaire fala en especial de Madame de Chatelet e de Federico de Prusia, pero o efecto total bascula sobre o grotesco. [CORGA: GI1996-01/47]

A esencia fica intacta, por moito que os voluntarios de hoxe en día, ademais de voluntarios, sexan internautas. [CORGA: CC2005-03/42]

O escritor Benedetti, que se encontraba onte en Madrid, é, ademais de poeta, narrador, dramaturgo, ensaísta, xornalista e traductor e viaxou por numerosos países europeos e iberoamericanos, así como por Estados Unidos. [CORGA: CG1999-06-01/44]

O evento foi transmitido por vídeo-conferencia nun total de 27 cidades en Brasil, ademais de puntos en Portugal, Paraguai e Galiza. [CORGA: CC2008-10/26]

• Parece ter natureza parentética.

• Semella actuar no interior da cláusula como primeira ou segunda parte dunha función sintáctica clausal case sempre descontinua. Nos seguintes exemplos vémolos no complemento de réxime da primeira secuencia e no atributo da segunda:

Ademais de si mesmo _o que fai con notable ecuanimidade_ Voltaire fala en especial de Madame de Chatelet e de Federico de Prusia, pero o efecto total bascula sobre o grotesco. [CORGA: GI1996-01/47]

A esencia fica intacta, por moito que os voluntarios de hoxe en día, ademais de voluntarios, sexan internautas. [CORGA: CC2005-03/42]

En suma, parece establecerse unha relación entre estruturas funcionais paralelas que respectan a función sintáctica do sintagma que acompañan, de xeito que se o

segmento análogo se constrúe con preposición, esta repítese no segmento con *ademais de*. Visualízase claramente o paralelismo nos enunciados que seguen mediante o subliñado co que poñemos en relación os dous membros da estrutura:

Ademais de insoslaible, dixo que a Diferencia é excluínte. [CORGA: C. Gonsar, *Arredor do non*, Edicións do Cumio, 1995]

O acordo base será asinado, ademais de polos traballadores, por Enrique Tatay e a inmobiliaria madrileña TAU, que comprará os terreos de Cabral. [CORGAetq: *CG1996-02-17/39*]

As eleccións, xa ven, serven para estas cousas, ademais de para elixir os nosos representantes políticos e darlles traballo ás imprentas. [CORGA: *CG1996-02-23/8*]

4.4.14.4.4. Debido a

No seu momento cuestionámonos se en XIADA o encabezamento deste apartado deberíamos consideralo locución prepositiva ou, pola contra, manter para el a análise desagregada correspondente ao participio de *deber* –[V0p0ms deber]– e mais á preposición.

Por unha banda, a conmutación natural pola preposición *por*, apoiada pola existencia doutras locucións prepositivas de significado tamén causal nas listaxes das *Normas* (RAG/ILG, 1995: 179-180; 2003: 180-181) (*a causa de, por mor de, por culpa de...*), facíanos pensar que lle deberíamos dar ese tratamento. Mais, por outro lado, a súa falta precisamente nas listaxes de preposicións dos manuais máis empregados (RAG/ILG, 1995 e 2003; Álvarez, Regueira e Monteagudo, 1986; Freixeiro, 2000; Álvarez e Xove, 2002 ou Hermida, 2004) facíanos dubidar. Decidimos, entón, que o máis efectivo sería a consulta ao CORGA co fin de comprobar a súa cohesión e invariabilidade.

Polo que respecta á primeira, constatamos que non estamos ante unha unidade cohesionada, posto que é posible intercalar no seu interior, por exemplo, adverbios, ben simples ben multipalabra:

En Estados Unidos, o descenso das vendas polo miúdo nun 0,3 por cento o pasado mes de xaneiro debido en parte á baixa demanda de automóbiles_ veu a pór de relevo a debilidade da economía norteamericana. [CORGA: *CG1996-02-28/1*]

O maior peso relativo dos ingresos reside nas transferencias correntes (61 por cento), pero o incremento máis espectacular produciuse no apartado de impostos directos, no que a

recadación creceu en máis dun 300 por cento (46.849 millóns), debido fundamentalmente á cesión do 30 por cento do IRPF ás comunidades autónomas. [CORGA: CG1999-05-22/46]

Na conta de resultados, Bankpime presentaba un incremento importante en tódalas súas magnitudes, desde os intereses e os rendementos, que pasaron de 8.204 no 1994 a 9.640 no 1995, debido especialmente ó rendimento da renda fixa que pasou de 2.273 millóns a 4.076 millóns, e tamén ás comisións que pasaron de 867 no 1994 a 1.060 no 1995, e en beneficios extraordinarios contabilizou Bankpime 1.567 millóns no 1994 mentres que no 1995 foron de só 466 millóns de pesetas. [CORGA: CG1996-02-25/80]

Por tipos de crédito, destaca a expansión de financiamento hipotecario, que subiu un 12,1 por cento en outubro de 1995 en relación cos doce meses anteriores, inda que esta porcentaxe se reduciu nos últimos meses debido, en parte, ó esgotamento dos recursos do Plan de Vivenda. [CORGA: CG1996-02-26/64]

Por outra banda, observamos ocorrencias nas que a suposta locución non está gramaticalizada, como nos demostra a concordancia:

Sen embargo, a medida que se achegaba a media sesión comezou unha clara tendencia á baixa, en parte debida á toma de beneficios logo das fortes alzas das tres últimas sesións. [CORGA: CG2000-11-02/11]

Un dos problemas máis salientables hoxe en día no eido da epidemioloxía da tuberculose, é a posibilidade da aparición e difusión de enfermidade debida a bacilos mono ou poli resistentes á terapéutica habitual. [CORGA: J. J. Gestal Otero, *Enfermedades infecciosas emerxentes*, Xunta de Galicia, 1997]

Imaxinacións túas, debidas a esa nervosidade que te domina. [CORGA: X. M. Martínez Oca, *Beiramar*, Xerais, 1983]

A auditoría do Consello chegará ó Parlamento nas próximas semanas marcada polo retraso de anos e importantes lagoas informativas debidas á inconcreción das contiosas partidas de gasto. [CORGA: TN1998-04/12]

fronte a estes outros:

As garantías estatais outorgadas ás compañías aéreas debido ó sobrecusto das primas de seguros como consecuencia do risco de ataque terrorista e guerra serán renovadas por un mes máis, con posibilidade de renovación tras este período. [CORGA: CG2001-10-17/25]

Ademais, Folgado indicou que no 2001 se producirá unha "flexión á baixa" da inflación en España, debido á esperada diminución dos prezos do cru e ó mellor comportamento dos prezos dos alimentos. [CORGA: CG2000-11-01/60]

Os dous factores anteriores, a intercalación de elementos no interior e as concordancias manifestas, lévannos a pensar que non estamos ante unha locución, razón pola que procedemos a realizar a análise desagregada: participio e preposición. Naturalmente, esta solución non empece que o conxunto sinala unha relación lóxica de tipo causal, como pode indicar a preposición *por* ou a locución *por culpa de*, mais non se debe a un cambio de significado de *debido* xunto coa preposición *a*, senón que é propio da construción pronominal que rexe complemento preposicional con *a*: *deberse a* = *estar causado por*.

4.4.14.4.5. Fóra de

Non queremos aquí volver sobre o xa dito para o grupo de adverbios nos que se inclúe *fóra*, para o que exemplificamos o seu tratamento un pouco máis arriba na epígrafe 4.4.5.4.1. cando tratamos o adverbio *dentro* seguido por un complemento coa preposición *de*. A nosa intención é outra: aclarar que a locución preposicional *fóra de* que aparece en XIADA é a que equivale a *excepto* ou *salvo*.

Entre as partículas exceptivas ás que lles prestamos unha especial atención, tanto aquí neste traballo coma noutro máis completo (Domínguez, 2012), non figura *fóra* nin *fóra de*, nin tampouco outras que se englobarían aí. O motivo non foi outro máis ca acoutar o terreo de investigación. De todas formas, dado que o obxectivo é só describir as peculiaridades da etiquetaxe practicada no CORGAetq, tampouco agora podemos afondar no seu estudo nin tampouco se entendería que aplicásemos na etiquetaxe dun corpus, que pretende ser de referencia, os resultados dunha teoría, por máis que creamos acertada, que escapa á caracterización predominante na lingüística galega. Así pois, dado que *a priori* nada nos fai desconfiar dunha posible delimitación incorrecta para *fóra* e *fóra de*, mantemos na clase das preposicións estes dous elementos nos usos paralelos a *excepto* e *salvo*.

A diferenza entre os dous pares (o adverbio, *fóra da casa* e a preposición ou a locución prepositiva *fóra/fóra de ti, non foi ninguén*) estriba en que o adverbio sinala unha referencia espacial, mentres que os elementos prepositivos indican a exclusión ou inclusión, dependendo da polaridade, dalgún elemento en relación cun grupo. O primeiro é antónimo de *dentro* e os segundos, equivalentes a *excepto*.

Dos exemplos seguintes, os dous primeiros ilustran a anotación no CORGAetq de adverbio modificado por unha frase preposicional con *de*, mentres que os outros catro representan a correspondente a elementos prepositivos:

Deste modo, a Caixa finalizará o exercicio cunha rede de 719 oficinas, das cales un 32 por cento estarán fóra de Galicia, precisou. [CORGAetq: CG2000-11-26/1]

Nesta nova edición, o autor decidiu engadirlle dous contos máis: Un máxico negocio e Un doce emprego, que foron escritos nas mesmas datas que os outros mais ficaron fóra da primeira edición. [CORGAetq: ANT2005-03-03/46]

Axiña has coñece-los meus fillos, son estupendos, fóra os seus pequenos defectos, quen non os ten. [CORGAetq: Ú. Heinze, *Remuíños en coiro*, Xerais, 1984]

Xa que logo, fóra de Cantabria, A Rioxa, Estremadura e a Rexión de Murcia, a canle xa está presente en todo o territorio español. [CORGAetq: GH2007-03-30/95]

Fóra da imaxe de feble saúde, a secuestrada falou varias horas e iniciou decontado unha xira que a levou a Francia e que pronto a deixará ver en Madrid ou Roma, por exemplo. [CORGAetq: ANT2008-07-10/82]

O Deportivo, coa formación que disputou o partido da Intertoto, fóra da inclusión do gardameta Daniel Aranzubia, tomou a iniciativa desde o primeiro minuto a través de rápidas e constantes combinacións entre Lafita, Juan Carlos Valerón e Sergio González. [CORGAetq: GH2008-08-05/76]

4.4.15. *Excepto, menos, quitado, quitando, sacado, sacando, salvo, senón*: problemas de categorización e tratamento en XIADA

4.4.15.1. Clasificación nos principais manuais actuais do galego

Álvarez, Regueira e Monteagudo (1986: 504, 505, 510, 511) caracterizan como preposicións *excepto, menos, quitado/quitando, sacado/sacando, salvo* e *senón*. Coa excepción de *salvo que*, a cal non recollen, consideran que as anteriores, seguidas da conxunción *que*, son locucións conxuntivas de tipo adversativo.

Costa Casas et al. (1988: 257-258) inclúen *salvo, excepto* e *menos* na relación de elementos que sen ser propiamente preposicións cumpren unha función relacionante similar ás que o son.

Freixeiro (2000: 543) engade á relación anterior *senón* entre os elementos que sen ser propiamente preposicións son capaces de cumpriren unha función relacionante do mesmo tipo que aquelas, establecendo a diferenza entre unhas e outras mediante

dous criterios: a) o caso do pronome que rexe cada unha, as propias esixen oblicuo fronte ás impropias que o piden recto, e b) a tonicidade, as propias son átonas mentres que as outras son tónicas. Non considera, non obstante, este autor que ningunha delas seguida de *que* constituía locución conxuntiva, pois non as recolle nin nas adversativas nin nas condicionais. Porén si recolle *só que* nas adversativas.

As Normas (RAG/ILG, 1995: 179-180; 2003: 180-181) recollen na nómina de elementos prepositivos, sen ningún tipo de aclaración posterior, *excepto, menos, quitando, sacado, sacando, salvo e senón*, ao lado de, por exemplo, *a, de* ou *sen*. Sorprende, non obstante, a omisión de *quitado* nesa listaxe mais é probable que se deba a un erro xa arrastrado dende a edición de 1995, pois séguese mantendo a paralela *sacado* e no DRAG (RAG, 1997) si figura como preposición²¹⁷.

Nas Normas (RAG/ILG, 1995: 182-185; 2003: 183-185) tamén entenden que, coa excepción de *menos* e *senón*, calquera das anteriores preposicións seguidas polas conxuncións *que* e *se* conforman locucións conxuntivas de tipo condicional: *excepto que, excepto se, quitado que, quitado se, quitando que, quitando se, sacado que, sacado se, sacando que, sacando se, salvo que, salvo se*. A variante con *que*, así mesmo, repite función conxuntiva no elenco de nexos de tipo adversativo cunha lixeira variación: engádense *menos que* e retírase *salvo que*. Non obstante, si consideran locución conxuntiva adversativa *só que*.

Tamén Hermida (2004: 189) inclúe na listaxe de preposicións impropias que denotan excepción, ademais de outras, *excepto, menos, quitado, quitando, sacado, sacando, salvo e senón*. Considera, así mesmo, que seguidas de *que* conforman locucións conxuntivas de tipo condicional. Non obstante, se as segue *se* non as trata como locucións senón como preposicións, que igual que as propias, poden preceder a conxunción condicional prototípica (2004: 213).

Álvarez e Xove (2002) desmárcanse da tónica xeral rexistrada nos distintos manuais galegos e clasifican as unidades en cuestión como conxuncións restritivas que introducen un tipo particular de oracións condicionais:

²¹⁷ En consulta realizada ao Seminario de Lexicografía o 28/03/2012, confirmáronos que tamén aparece na edición prevista para o 2012 e confirmámolo posteriormente nós mesmos cando neses días se puxo accesible a nova versión no enderezo <http://www.realacademiagallega.org/diccionario>.

Un tipo particular de oracións condicionais está iniciado por sintagmas constituídos por unha conxunción restrictiva (equivalente a 'pero non') e as conx. *se* ou *que* (*aga ~ agás se / que, bardante que, excepto se / que, quitado ~ quitando se / que, sacado ~ sacando se / que, salvo se / que, senón que*) ou outros sintagmas como *a menos que, non sendo se / que, a non ser se / que, de non ser se / que* e *catar se*. Nestas oracións a complexidade sintáctica é maior ca nas outras oracións condicionais, pois o termo que antecede a *se / que* (*excepto, non sendo...*) introduce un reparo ou contraposición de signo contrario ó cumprimento da aserción feita na principal, coma unha sorte de adversativa da anterior e de prótase reducida. A oración conxuncional con *se / que* establece as condicións en que se cumpre a restricción contida no termo anterior. As oracións de *que* constrúense con Subx. e as de *se* con Ind. ou Subx. Pto.[...]

Con algúns destes nexos tamén é posible o verbo en Xer. e en Part., mais en tal caso sen o elemento conxuntivo final:

Non os vas quitar de enriba, *salvo dándolles dunha vez o que che piden*.

Non lle vexo sitio apropiado, *excepto posto no final do corredor*. [Álvarez e Xove, 2002:158-159]

Cunha e Cintra (1984: 552) recollen tamén para o portugués entre as preposicións *acidentais excepto, menos, salvo* e *senão*; Vilela (1999: 250-251), en cambio, só menciona en distintos momentos como elementos prepositivos *excepto, salvo* e *menos*, mais a súa pretensión non é proporcionar unha nómina exhaustiva senón só amosar as características que as definen: procedencia, significado, polisemia, etc.

4.4.15.2. Breve revisión sobre os problemas para a delimitación categorial

Salvo Álvarez e Xove (2002) quen as consideran conxuncións, a xeneralidade das gramáticas e manuais actuais existentes para o galego tratan *excepto, menos, salvo* e *senón* no paradigma das preposicións, igual ca as dúas gramáticas máis coñecidas do portugués (Cunha e Cintra, 1984 e Vilela, 1999). Malia que recoñecendo a problemática da súa clasificación e exceptuando *sino*, na tradición hispánica consideráronse maioritariamente preposicións (Alarcos, 1994; DRAE, 1992 ou Pavón Lucero, 1999) ata que na *Nueva Gramática de la lengua española*²¹⁸ (2009: 2463-2468), tras repasar unha serie de propostas diferentes segundo as cales as partículas exceptivas poderían clasificarse como preposicións, adverbios, conxuncións subordinantes e conxuncións coordinantes, decántanse finalmente pola súa categorización como conxuncións; de feito, aparecen xa baixo a epígrafe “Conjunciones que denotan excepción”.

²¹⁸ En diante, NGLE.

A continuación imos resumir moi brevemente as razóns que se aducen na NGLLE para desbotar a elección de cada unha das clasificacións anteriores para *salvo*, *excepto* e *menos*, posto que todas elas son de aplicación no caso galego.

Así, en contra da súa delimitación na clase das preposicións están os seguintes factores:

- *Salvo*, *excepto* e *menos* poden ir seguidos de xerundio e entre as preposicións só *en* pode facelo.
- As secuencias de preposicións están severamente restrinxidas por factores semánticos, mentres que as partículas exceptivas poden ir seguidas de calquera preposición.
- As partículas exceptivas rexeitan as construcións relativas.
- Non admiten pronomes en caso oblicuo.

En contra da súa clasificación entre as conxuncións:

- *Salvo*, *excepto* e *menos* rexeitan os verbos en forma personal.
- *Salvo*, *excepto* e *menos* admiten infinitivos e xerundios.
- *Salvo*, *excepto* e *menos* preceden os dous termos que unen, propiedade que non posúen as conxuncións coordinantes.

Finalmente, repercuten negativamente na súa posible categorización na clase dos adverbios os seguintes:

- *Salvo*, *excepto* e *menos* non poden incidir sobre verbos en forma persoal, fronte aos adverbios.
- Os adverbios de inclusión poden seguir os elementos que modifican, mentres que as partículas exceptivas non, e ademais non hai que poñelos en relación cun antecedente.

Así pois, parece que non acaban de encaixar ben en ningunha das clases gramaticais existentes, xa que en contra de todas elas existen razóns de peso.

4.4.15.3. O seu tratamento en XIADA

Non é este o lugar para profundar en cada un dos argumentos acabados de relacionar nin tampouco o obxectivo desta tese. A nosa pretensión céntrase só en amosar a problemática que rodea estas partículas e como, téndoa en mente, procedemos á súa etiquetaxe en XIADA.

4.4.15.3.1. Excepto, menos, salvo

No caso de *salvo* e *excepto* optamos sempre pola delimitación na categoría das preposicións, por ser esa a categorización predominante na lingüística galega. Para *menos* tamén remitimos á clase Preposición os usos paralelos a *excepto* e *salvo*, integrando os demais na de Adverbio. Vexamos algún exemplo para cada un deles:

Excepto:

O Congreso apoiou todas as emendas do Senado, excepto a que impoñía a condición de ser doutor para que un profesor permanente puidese ser elixido decano de facultade ou director de escola universitaria. [CORGAetq: GH2007-03-30/26]

Excepto nos 15 días da campaña, é moi difícil ver os políticos galegos pisando corredoiras, lonxe do seu hábitat natural de asfalto e moqueta. [CORGAetq: DLV2007-05-15/22]

O Concello da Coruña recibiu o visto e prace de Política Territorial para a devandita modificación ao considerar adecuada toda a ordenación aprobada provisionalmente polo Concello, excepto para unha superficie de 4.007,04 metros cadrados de titularidade municipal que supón un 1,16% da superficie total de 344.747 metros cadrados. [CORGAetq: GH2008-08-05/17]

Menos:

Sen embargo, o horizonte é agora calquera cousa menos aséptico. [CORGAetq: ANT2005-03-03/66]

Así, mentres Sanidade afirma que só un 12% de cirurxías foron aprazadas, a portavoz da Asociación de Facultativos especialistas Interinos nos Hospitais do Sergas (Afeihs) afirmou que "se suspenderon practicamente todas as operacións programadas" menos as de urxencia. [CORGAetq: GH2007-03-30/8]

Todos, menos Mary Kate, que rexeitou falar", díxolle ao diario unha fonte relacionada co caso, que tamén explicou que Olsen podería ser "a testemuña que se necesita para concluír a investigación sobre quen lle proporcionou as drogas e os medicamentos". [CORGAetq: GH2008-08-05/48]

Salvo:

As entidades financeiras galegas debuxaron conxuntamente unha liña ascendente, salvo excepcións puntuais coma o Banco Simeón, que en 1995 pasou a mans da Caixa Geral de Depósitos, ou o Banco Etcheverría, cun retroceso que non afecta a suma global. [CORGAetq: CG1996-02-15/33]

Un resultado non despreziable de todo isto é que o número e gravidade de conflitos bélicos e atentados terroristas foi menor neste ano, salvo en casos especiais de enconamento histórico. [CORGAetq: ANT2002-09-12/31]

Por agora, as estratexias postas en marcha non están a dar os resultados apetecidos salvo a redución do paro debido á mellora das taxas de medre económico dos últimos anos pero, tamén, ás sucesivas modificacións na definición de parado e á perda de calidade e veracidade das estatísticas oficiais. [CORGAetq: ANT2002-10-10/32]

A maiores, integramos nos usos prepositivos de *menos* os relacionados con calquera expresión do signo “-” en linguaxe matemática, por crer que o significado que posúe a preposición exceptiva é o que se aplica en calquera representación con números, incluída a expresión horaria:

Mentira, é a unha menos cuarto. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Doce menos cinco. [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

En XIADA non consideramos que *salvo*, *excepto* e *menos*, seguidas da conxunción *que* ou *se*, cando a súa significación é a de excepcionalidade ou condicionalidade, conformen locucións conxuntivas por canto se explicou na epígrafe 4.4.5.2.1. e mais por canto as probas que acostuman empregarse para o recoñecemento das locucións –significado diferente do da suma dos elementos que a compoñen, rección modal e grao de gramaticalización– nos indican que non o son. Así, en calquera desas aparicións no noso sistema, trátanse como unha preposición calquera seguida dunha conxunción subordinante. Por exemplo:

De tanto chover, por non haber xa non hai soberanos en Grecia, nin atrancos (salvo que Don Juan Carlos fale de Irlanda)... [CORGAetq: DLV2007-05-15/39]

Calquera Observador Alleo á Movidá pensaría que todo aquilo era normal, salvo que se lle fíxese ver esa coincidencia: as caras de todos eles, as cabeleiras canosas, o estilo case Policial de fumar e de fugar na orella coa deda mediante series prolongadísimas de sacudidas coma unha vibración de man e antebrazo, a indumentaria cutre a base de cazadoras gastas e pantalóns

anchos de tergal, as olladas de Inexplicable Cansazo e as pálpebras color merda. [CORGAetq: X. Cid Cabido, *Días contados*, Xerais, 1991]

Polo momento non tivemos máis casos ca os anteriores. *Excepto e menos* non concorren no que levamos revisado nin seguidas de *que* nin de *se*. Si hai algunha mostra no CORGA, mais non son moi frecuentes. En todo caso, a intercalación dalgún elemento entre os dous constituíntes da suposta locución, como nos demostra o último dos seguintes exemplos, indica que a cohesión interna da posible locución é fráxil:

"Non teño nada que dicir contra os asturianos, excepto que son agarrados e mesquiños" [CORGA: *TN1997-10/2*]

Sobretudo as últimas, as definitivas palabras con que concluía o romance: nunca o rouxinol quererá esquecer o canto, excepto se este é roto polo acróstico paulatino, como na infancia. [CORGA: R. Raña Lama, *Nun cuarto para agardar / O último hexagrama / Cantar sen xesta de Lanzarote do Lago / Destino de Lumia / Grabación da volta a Galiza, dende o pazo de Raxoi*, Ediciós do Castro, 1985]

Contoume que el, ao contrario ca min, non prefería a compañía dos homes, salvo se era inevitábel. [CORGA: M. Lourenzo González, *Noites de papel*, Xerais, 1997]

Un contexto no que, desgraciadamente, a sociedade no seu conxunto ten pouco que gañar, e moito que arriscar. Salvo, claro está, que funcionen axeitadamente as demais institucións indutoras de estabilidade nos sistemas democráticos: nomeadamente a xustiza, e particularmente a xustiza constitucional. [CORGA: *GH2007-03-11/74*]

Tema distinto é a combinación de *excepto*, *menos* e *salvo* con *que* en modo indicativo, do que non temos como xa dixemos ningún caso aínda no CORGAetq, cando non exclúe ningún elemento dun grupo dado nin expresa condicionalidade, senón simplemente matiza ou introduce unha corrección sobre o expresado anteriormente. Nestes casos non nos vale a conmutación que propoñen Álvarez e Xove (2002: 158) por “pero non” e a combinación semella comportarse como unha verdadeira locución, equivalente a *só que* ou *pero*:

Todo estaba ben excepto que saíamos das duchas totalmente enxabonados, cheos de espuma que non podíamos quitar. [CORGA: C. Franco, *A lúa no cénit e outros textos*, Ediciós do Castro, 1988]

Os ollos de Helena tiñan un aspecto moi semellante ós poemas de Francisco de Sás, excepto que ó pasa-las follas apareceu un debuxo no comezo dunha delas, unha raíña ou princesa vestida cun traxe vermello e cuberta cun manto azul, orlada a súa figura de pequenas volutas de ouro. [CORGA: M. Aleixandre, *Tránsito dos gramáticos*, Xerais, 1993]

4.4.15.3.2. *Senón, senón que*

O caso de *senón* é bastante máis complexo por canto, en teoría, temos un *senón* preposición, un *senón* conxunción adversativa e un *senón que* locución conxuntiva de tipo adversativo e condicional. Álvarez e Xove (2002) e Hermida (2001, 2004) inclúen tamén nos usos conxuntivos o equivalente a *máis ca* e o copulativo.

No sistema de XIADA só temos recollidas a análise como preposición, da que non detectamos ningún caso aínda no corpus de adestramento, e mais a análise como conxunción subordinante, que aglutina tanto os usos adversativos (*Non teño sede senón fame*) coma os aditivos (*Non só teño sede senón tamén fame*) e os exclusivos (*Non teño senón fame*), ademais da locución conxuntiva subordinante *senón que*.

Exemplos do seu uso preposicional cun valor semellante ao de *excepto, menos e salvo* –*Senón eu, sabíano todos*–, son os seguintes:

Mais se un palmo se arredaren do río, se por temeridade ou ousadía os homes de pelo roxo se adentraren fóra destas ribeiras, a nosa palabra a nada nos obrigarán, e quedaremos libres de mantela, e ninguén nos pedirá contas senón da vitoria que logremos. [CORGA: V. F. Freixanes, *A cidade dos Césares*, Xerais, 1993]

Semellaba moi afectada pola morte do camioneiro da que ninguén senón ela, parecía responsabilizala. [CORGA: C. Vila Sexto, *Cara á fin da luz*, Xerais, 1995]

A noite viña novamente ó meu encontro e ninguén senón aquel mastro esguío dera sinal de presenza. [CORGA: X. Queipo, *O paso do noroeste*, Sotelo Blanco, 1996]

¿Quen ía hoxe ocupar tal cargo, senón eu que fago de todo, e aínda me faltaría un irmán máis... [CORGA: A. Álvarez Cáccamo, *O espírito de Broustenac*, Xerais, 1996]

Entrevistador: _Malia que a propaganda fraguista diga que Galicia ten peso no exterior desde que goberna Fraga, case sempre está ausente nas grandes operacións de Estado, e ninguén a cotexa xunto a Catalunya e Euskadi senón formando paquete coas rexións. [CORGA: *TN1997-10/11*]

Hoxe en día parece que *menos*, e en menor medida *salvo* e *excepto*, desprazou *senón* en contextos positivos para expresar a excepcionalidade. Na versión 1.5 do CORGA non demos con ningún exemplo; no CTAG tampouco e no TILG, curiosamente etiquetada como conxunción (valor numérico 681), realizando a consulta que abrangue todo o período temporal alí tratado sobre o lema *todo* (termo de

polaridade positiva) que debía estar seguido polo lema *senón* a unha distancia de como moito catro palabras²¹⁹ só aparece o caso seguinte:

Remedio hai ño pra todo , señorita , senón pra a morte . [TILG: M. Valladares, *Majina ou a filla espúrea*, La Ilustración Gallega y Asturiana, 1880]

Non obstante, como indica por exemplo Muñoz de la Fuente (1996), era unha forma máis de expresar a excepción na época medieval, xunto con *salvo*, *fora*, *menos*, *salvante*, etc.

O *senón* adversativo, complementario de *pero* e *mais*, relaciona dous elementos (cláusulas, frases ou palabras) entre os cales o primeiro expresa negación e o segundo, o que introduce *senón*, unha noción afirmativa. Mostra destas análises en XIADA son as seguintes secuencias:

O concepto da economía no inmediato futuro non se vai centrar tanto en crecer, senón en como crecer. [CORGAetq: *GII1996-06/30*]

Bugallo apuntou que Cultura non dará créditos senón subvencións. [CORGAetq: *GH2007-03-17/44*]

Non queriamos reducir o galeguismo a un partido senón galeguizar todas as ideoloxías e os rapaces irían uns ao cristianismo, outros ao socialismo... [CORGAetq: *ANT2008-07-10/22*]

En primeiro lugar, non son movementos sociais senón montaxes dende centros de poder políticos, disfrazados de movementos sociais. [CORGAetq: *GH2008-08-05/43*]

En correlación con *non só*, *non soamente*, *non exclusivamente*, *non unicamente* deixa de ter valor adversativo (non A si B) para amosar o de adición ou copulativo (A e tamén B ou nin A nin tampouco B)²²⁰. Acostuma ir acompañado por adverbios como *tamén*, *ademais* ou *incluso* que, igual ca o adverbio *só* ou os xa citados en *-mente*, reforzan a negativa e inciden significativamente na adición. Igual ca calquera conxunción copulativa pode unir distintos tipos de segmentos: palabras, frases ou cláusulas, mais, a diferenza dos nexos propiamente copulativos, non pode unir máis ca dous elementos. Vexamos unha pequena mostra:

²¹⁹ Desafortunadamente, nin o sistema de consultas do TILG da USC [<http://www.ti.usc.es/TILG/>] nin o do RILGA [<http://sli.uvigo.es/TILG/>] permiten o emprego de operadores booleanos nin que sexa o usuario quen determine a distancia á que poden estar situadas as palabras obxecto da busca.

²²⁰ Sería interesante estudar se este uso non é unha mera variante enfática da construción exceptiva ou exclusiva, o que impediría que poidan sumarse máis de dous elementos por similitude con aquelas.

Por paradóxico que pareza, a crise non nos ofrece so a cara do empobrecemento, senón tamén a fasquia dunha singular oportunidade para Galiza iniciar un desenvolvemento autocentrado, do que careceu astra hoxe. [CORGAetq: MC1980-12/1]

Con todo, como dicían os vellos, este mes non é de fiar, non xa por cambiar dun día para o outro, senón mesmo ao longo da xornada: [CORGAetq: GH2007-03-17/79]

Non só os grupos parlamentarios da oposición _entón BNG e PSdeG_ puxeron en dúbida o custe da celebración do Ano Santo, senón tamén destacados representantes da cultura e da intelectualidade galegas. [CORGAetq: TN1998-04/12]

Cremos que é importante destacar que hoxe en día *senón*, a diferenza dos outros nexos adversativos prototípicos *pero* e *mais*, necesita invariablemente da conxunción *que* para poder combinarse con formas verbais flexionadas, tanto cando se contrapón a un elemento negativo como cando é de adición. Malia que en ambos os casos debера poder unir directamente cláusulas, non atopamos máis ca un exemplo no CORGAetq e outros dous no CORGA:

E logo, Castelao porque os aspectos da visión económica que están nas súas obras, fundamentalmente en Sempre en Galiza, que non son só seus senón tamén están as cabezas do nacionalismo galego do primeiro terzo do século vinte do partido galeguista, cunha diagnose da estrutura da sociedade. [CORGAetq: GH2008-08-05/43]

Contéplase unha reunión extraordinaria en Lisboa, á que están convidados os países africanos dos que proveñen os inmigrantes para pór en marcha mecanismos de cooperación dende unha visión máis global, na que non só se aborde a "bunkerización" europea, ou as necesidades do seu mercado laboral, senón se habiliten recursos e políticas que incidan nas causas reais e os efectos para países emisores e receptores do fenómeno migratorio. [CORGA: ANT2007-02-22/60]

Noutros casos o motivo non é a falta de cultura senón é que a balanza da vide inclínouse do lado da produción e non do da calidade, entendida como autenticidade, personalidade... [CORGA: ANT2007-02-22/70]

Finalmente, o outro uso conxuntivo recoñecido en XIADA intégrase no que Álvarez e Xove (2002) ou Hermida (2004) denominan usos similares a *máis que / ca*, posto que é posible conmutalo por esa combinación. Non obstante, debemos recoñecer que na nosa opinión este uso está próximo ao de *salvo* e *excepto* ao introducir un elemento que se sinala como único fronte ás demais opcións, que se negan implícita ou explicitamente no resto da oración, e que sería necesario afondar moito máis no estudo

deste *senón*. Mostra etiquetada como conxunción subordinante en XIADA desta significación son as ocorrencias seguintes:

En primeiro lugar, a medida que avanza no libro, o lector comeza a sospeitar que a escritura en viñetas non é senón un recurso destinado a ocultar a incapacidade para facer progresar unha historia, para crear personaxes cun mínimo de coherencia, para dotar de sentido a narración. [CORGAetq: ANT2008-07-10/30]

De feito, a sex simbol que marcou os anos 60 e 70, alcumada O Corpo, non era senón Raquelita Tejada, filla dun inmigrante boliviano e emparentada coa que foi presidenta de Bolivia Lidia Gueiler Tejada. [CORGAetq: ANT2008-07-10/54]

Por facilitarlle o labor ao etiquetador e aos usuarios do corpus, decidimos tratar sempre a combinación *senón que* como locución conxuntiva subordinante, de modo que sexa fácil recuperar os seus contextos de uso e poder logo estudalos en profundidade. Así, caen baixo esta etiqueta as ocorrencias propiamente adversativas nas que se dá unha contraposición na que o membro introducido por *senón* achega a parte positiva (exemplos 1 a 3), e mais as aditivas, nas que non se produce contraposición entre dous elementos *senón* a suma de ambos (exemplos 4 a 6):

1. "Somos gaiteiros do século XXI, non podemos vivir cunha ollada retrospectiva, senón que debemos utilizar esta mirada para desenvolvernos neste tempo como músicos actuais". [CORGAetq: ANT2005-03-03/69]
2. Unha opinión mantida por gran parte dos musulmáns, que pensan que o monarca saudí "non representa ao Islam, senón que fai dano á memoria do Profeta e aos auténticos crentes". [CORGAetq: ANT2008-07-10/10]
3. É previsíbel que este anuncio non achegue medidas moi diferentes ás do "Plan de Acción 2008-2012", presentado en xullo de 2007, senón que afonde nas que alí se establecían. [CORGAetq: ANT2008-07-10/1]
4. A muller non só non sorri senón que manifesta ás claras a súa molestia, e antes de que o home abra a boca, inquire: [CORGAetq: A. Álvarez Cáccamo, *Catapulta*, Xerais, 1995]
5. A portavoz da Coordinadora de ONG's referiuse tamén ó caso da Asociación Valenciana de Cooperación Iberoamericana, da que di que "non só non sabemos se existe ou non, senón que o Consell concedeulle esta na súa reunión do 29 de decembro, polo que é difícil saber se o trámite administrativo que se utilizou foi o correcto". [CORGAetq: CG1996-02-29/104]
6. Fontes de Grummar explicaron que estes días está tamén pola zona unha manda duns 35 exemplares de golfiño, pero non só non se sumou a eles, senón que tivo varios encontros pouco amigables. [CORGAetq: GH2008-08-05/41]

No corpus de adestramento non demos aínda con ningún exemplo mais si constatamos tamén no CORGA o valor condicional (exemplos 7 a 9) e o correctivo (10 a 12), que serán etiquetados tamén como locucións de tipo subordinante sen discriminacións ulteriores:

7. Ricardo: E así os que caen sempre son poucos, que eles non se comprometen senón que sexa sobre seguro. [CORGA: M. Riveiro Loureiro, *O coronel ten a quen lle escribir*, Xerais, 1989]
8. Felipe: Pois xa non é cousa súa o que aquí poida acontecer, senón que queira estorbar agora onde antes non axudou. [CORGA: M. Riveiro Loureiro, *O prazo*, Xerais, 1994]
9. O prometedor concelleiro botouse con grande entusiasmo á fundación dun partido político disque independente, coa máis formal independencia anque semelle moqueo, e segue na teima de lle dar solución ós problemas populares anque sexa a base de acabar co pobo, que ninguén sabe moi ben para que serve senón que sexa para amolar. [CORGA: M. Riveiro Loureiro, *Corpo canso*, Ir Indo, 1993]
10. A fonte estaba onde meu pai dicía, e aínda segue hoxe alí, senón que agora está toda encanada, e hai aparcadoiro para os coches. [CORGA: D. X. Cabana, *A excursión*, Transportes Castromil, 1992]
11. En conxunto, quedoulle un caseto bastante xeitoso, parecido a un cabazo dos da Terra Chá, senón que máis ancho. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]
12. Se ben se mira, sonche coma un mestre, senón que eu aprendo respecto e non contas. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Fóra todo o anterior, hai outros casos de *senón* que merecen un comentario á parte sobre como se tratan no noso sistema. Por unha banda, están aqueles en que se produce na escrita a amálgama incorrecta da conxunción condicional *se* e mais o adverbio de negación (*senón* en vez de *se non*). No corpus desambiguado só demos con un, pero no CORGA hai bastantes máis. Nestes casos mantemos en unidade o segmento tal e como nos aparece escrito, mais logo na etiquetaxe consideramos dous constituíntes, dos cales o primeiro é a conxunción subordinante *se* e o segundo o adverbio nuclear *non*:

Por qué hei de chorar por ninguén, dille, senón me deixarán chorar por min. [CORGAetq: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Pois ningunha hipótese resultaría cientificamente relevante senón é transcribible na linguaxe formal daquela ciencia á que pretenda incorporarse. [CORGA: A. Carreira Castro e C. Padilla Heredia, *Filosofía da Ciencia e da Tecnoloxía*, Bahía, 1998]

E é que certamente non se entendía a orde senón era como relación entre elementos parados, como inmovilidade. [CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]

Estas regras, por outra banda, non son posibles, senón partimos do coñecemento dos elementos que crean unha conducta e das leis dos mesmos. [CORGA: R. Regueira, *X.V. Viqueira: Teoría e praxe*, Novo Século, 1992]

No mundo non hai bárbaros porque si, non hai nin sequera persoas malvadas porque si, senón é porque as creamos; [CORGA: *CG2001-10-09/42*]

Senón se quere intervir directamente tamén se pode investir en empresas vinícolas que cotizan en bolsa. [CORGA: *ANT2002-10-17/29*]

Por outra banda existe un número bastante grande de ocorrencias no CORGA de *senón* que, entre outras características, non se constrúen con polaridade negativa, poden ir seguidas de forma verbal flexionada, aparecer na posición final, ir precedidas da conxunción *e* ou *pero*, iniciar secuencia... Por exemplo:

_Efectivamente, senón non podería montar o meu telescopio. [CORGA: A. Fernández Paz, *Avenida do Parque, 17*, SM, 1996]

Prot: Tampouco podía poñerme a razoar filosoficamente diante daquela xente, porque senón decatárianse de que eu era un intelectual, e preguntárianse o que fai un intelectual traballando de camareiro, porque un intelectual non traballaría nunca de camareiro, todo o mundo o sabe, ¿que intelectual limpa mesas? [CORGA: R. Salgueiro, *Todo é horrible, moi horrible e eu tamén*, Asociación Socio-Pedagóxica Galega, 1996]

A escola de prácticas de torreiros estaba aquí en Galicia, naturalmente, senón onde ía estar, coa cantidade de costa que aquí temos ben batida do mar, con tantos faros que aquí hai para practicar, e todos diferentes uns dos outros, ¿non é? [CORGA: M. Riveiro Loureiro, *O Patrón*, Edicións do Cumio, 1990]

A Meiga: ¿Por qué outra cousa ibades acudir a min senón? [CORGA: E. R. Ruibal, *O cabodano. Cousas da morte. A sombra do bon cabaleiro*, Pico Sacro, 1977]

Emperador: Máis che valerá que sexa algo importante, porque senón hei mandar que te metan na maseira na que comen os crocodilos. [CORGA: M. Núñez Singala, *O achado do castro*, Sotelo Blanco, 1992]

Catulo: Menos mal que aínda nos rimos algo, senón, una tarde perdida. [CORGA: M. Núñez Singala, *O achado do castro*, Sotelo Blanco, 1992]

Capitán Martín: ¿Onde senón, van ir dous soldados valentes? [CORGA: M. A. Murado, *A grande noite de Fiz*, Xerais, 1994]

En canto ao neboeiro, penso que temos moito mais del no Ribeiro, en Monforte ou en Pontearas que nas amplas rúas londinenses, e senón preguntémosllo aos sofridos viaxeiros de Peinador que en pleno mes de xuño e de setembro, viron os seus voos desviados a Santiago por mor do neboeiro porriñés. [CORGA: *ANT2002-10-31/74*]

Cando queres déitaste ó sol, e se non hai sol fas lume e pásache por vez, e non tes que estar preso das vacas, nin dun traballo que tes que ir a tal hora, e senón bótante e ó carallo... [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Agora mesmo miro se a hai na cociña, e senón baixo nun voo a comprala." [CORGA: X. M. Martínez Oca, *Náufragos en terra. Os amores imposibles*, Xerais, 1995]

Aínda estamos na xornada vinte e seis, aínda que si é un partido importante, polo que o ideal sería gañar, pero senón, polo menos, non perder, e con isto non quero dicir que non vaiamos pola vitoria, pero hai que pensar que necesitamos dezaseis puntos para salvarnos e todo o que sexa sumar, onde sexa, como sexa, é bo", sinalou Vázquez [CORGA: *GH2007-03-11/65*]

Aquel bellaco incapaz de sentimentos nin fidelidades ocultaba demasiado. Senón, ¿a que viña aquela rapidez?; [CORGA: X. R. Pena, *O reverso do espello*, Xerais, 1984]

Os manuais que recollen a correcta grafía de formas conflitivas coma *senón* (por exemplo, Hermida 2001: 63-64) indican que os casos dubidosos se presentan cando a oración condicional está elidida e a única marca que deixa é a conxunción e mais o adverbio *non*. Os casos que vimos de ver comparten todos esa característica: neles é posible recompoñer unha estrutura condicional con *se*. Agora ben, tamén é posible conmutar por *en caso contrario, do contrario*.

No TILG²²¹ remiten á clase gramatical Conxunción (numericamente a 681) estas ocorrencias de *senón*. Por exemplo:

Iste segundo , despois de ouvir ÷lle ó primeiro tantos eloxios , fartou ÷se e dixo ÷lle : ¡ Mira , a túa , porque che saliu boa , senón non a dominas nin ti nin Dios ! [TILG: X. Estévez, *A Lagoa da memoria*, 2002]

^boa/bo_063 ^/,_999 ^senón/senón_681 ^non/non_661 ^a/o_153 ^dominas/dominar_527
^nin/nin_681 ^ti/ti_119 ^nin/nin_681 ^Dios/deus_042 ^!/!_999 ^Mais/mais_681
^sorna/sorna_043 ^tiña/ter_541 ^iste/este_352

²²¹http://sli.uvigo.es/TILG/tilg_pesca.php?pesca=sen%F3n&tipo_busca=palabra&categoria=Todas&desde=1600&ate=2050&pesca2=non&posicion=1&tipo_contexto=palabra [consulta realizada o 8/4/2012]

Nós cremos que efectivamente a orixe deste *senón* que tanto pode encabezar a secuencia na que se insire como ocupar a posición final, e que non se contrapón a nada, pode estar na elisión de parte do CONDICIONANTE²²² dunha estrutura condicional. Non obstante, o feito de que nun rexistro coidado e escrito apareza con tanta frecuencia amalgamado²²³, ademais de non sempre esixir contexto positivo previo, indícanos que se produciu ou está a producir a gramaticalización desta unidade, malia que sexa arriscado afirmalo, nunha de tipo adverbial.

Aínda que na maioría dos casos basta con botar man da afirmación ou negación inmediatamente anterior para reconstruír o CONDICIONANTE –exemplos de b–, tamén sería factible reconstruír para unha estrutura exclusiva –exemplos de c–:

a) Daquelas palilladas sairon sempre as palilleiras e os mellores encaixes de bolillos. ¿Onde, senón, ían aprende-las que desexaban introducirse nos coñecementos da técnica? [CORGA: X. L. Blanco Campaña, *Camariñas terra de palilleiras*, Xerais, 1991]

Fiz: Pero estades vestido de capitán. / Capitán Martín: ¿E de que senón? Son capitán. [CORGA: M. A. Murado, *A grande noite de Fiz*, Xerais, 1994]

b) ¿Onde ían aprende-las que desexaban introducirse nos coñecementos da técnica se non era nas palilladas?

E se non me vestise de capitán de que me vestiría?

c) ¿Onde, senón nas palilladas, ían aprende-las que desexaban introducirse nos coñecementos da técnica?

¿E de que, senón de capitán?

Con todo, non sempre é só unha parte do contexto anterior a que nos axuda a interpretar o *senón*. Na seguinte secuencia precisamos de todo o anterior para interpretar correctamente. Non nos vale cambiar o *senón* por *se non a perdeu*, temos que á forza conmutar por *se a perdeu*, co que xa non estamos só ante unha cuestión de elisión:

Anque o que se di perder a idea que tanto o levaba non debeu perdela. Senón non se explica o feito de que trouxese o tallo acó e o pasease arreo por toda a vila. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

²²² Función que na corrente funcionalista se lle asigna ao membro que achega a conxunción e introduce a condición para que se cumpra o exposto no outro membro: *Se non chove*, non me vai medrar a horta..

²²³ Só seguido de *non*, *senón non*, supera os 50 exemplos.

No corpus de adestramento de XIADA preferimos ser conservadores e etiquetar como conxunción, xa que a súa categorización como adverbio implicaría, probablemente, que a información non se recuperase por non estar dispoñible aínda o manual de etiquetaxe.

Relacionados cos anteriores, pero nunha variante distinta, están os casos en que, coa aparencia dunha estrutura adversativa na que se contraponen dous termos, aparece, a primeira vista, a ANTÍTESE antes ca a TESE. Os manuais din que isto non é posible, pero como se explican logo casos coma os seguintes?

Ó fio do mediodía acordaron os visitantes e os cabaleiros, xuntamente con Sancho, considerado senón coma un deles si digno de acompañalos polo seu historial e porque a el tamén querían as monxas honralo, xantaron coa Madre Superiora, Dona Nazarina do Bordelo Ramalleira, madura e fermosa, aínda, dona de redondeadas formas, baixo as toucas, quen se congratulou moitísimo de telos alí e avisounos de que á tardiña tería a comunidade o pracer de coñecer, en persoa, a tan esforzados cabaleiro e[texto omitido] [CORGA: L. Vázquez Xil, *A da alba sería*, Edicións do Cumio, 1996]

[texto omitido]en primeiro lugar, que a súa designación, seguindo as pautas relativas á propia Mesa do Parlamento (artigo 41 do Regulamento), tende a respecta-la proporcionalidade de tal xeito que os postos de Presidente, Vicepresidente e Secretario, repártense negociadamente senón entre tódolos Grupos sí entre os máis numerosos da Cámara. [CORGA: A. Teixeira Piñeiro, *O Parlamento de Galicia*, Ir Indo, 1988]

Certamente eu chegara senón ó namoramento, e penso que si, sí ó prendamento do Quietán; [CORGA: A. Conde, *Memoria de Noa*, Xerais, 1982]

a cidade inda era un monstro durmido e podre con ben poucos vermes a percorre-lo que, os xornalistas, chaman arterias cunha impudicia absoluta e lamentable, pero que eiquí, non é que veña como anelo ó dedo, precisamente, mais semella, senón axeitado, si, cando menos, inevitable e ramplón; [CORGA: A. Conde, *Memoria de Noa*, Xerais, 1982]

No sistema XIADA tamén, de documentárense no corpus desambiguado manualmente, serán etiquetados como conxunción; mais, na nosa opinión, asistimos novamente á amalgama incorrecta da conxunción *se* e o adverbio de negación, pois cremos que o *non* forma bloque sintáctico co segmento que o segue e que é con este co que contrasta o posterior adverbio de afirmación *si*. Malia que non localizamos ningún caso no CORGA, sería posible, pospoñer o adverbio de negación (*se axeitado non, si inevitable...*), ou mesmo intercalar un adverbio (*se probablemente/quizais non axeitado*,

si inevitable...). O CORGA tamén achega exemplos para esta interpretación, velaí algúns:

restrinxida no seu despliegue social, abandonada ás súas propias forzas de lingua cuase exclusivamente oral, e alonxada dos centros de decisión política e económica, corre perigo, se non de morrer, si de deixar de ser a lingua maioritaria de Galicia. [CORGA: M. A. Araúxo Iglesias, *A fe cristiá ante a cuestión da lingua galega*, Edicións do Adro, 1975]

Estar dúas semanas lonxe do balbordo das políticas e das saídas ó paso era un bo motivo para sentirse, se non feliz, si tranquilo. [CORGA: R. Fonte, *As regras do xogo. Dez relatos en xénero negro*, Edicións do Cumio, 1990]

A principal dificultade que deberán afrontar os autores de linguas filosóficas radica, precisamente, en dar cunha disposición, se non tan racional como eles pretenden, cando menos si xuízosa dos entes. [CORGA: T. Moure, *Outro idioma é posible. Na procura dunha lingua para a humanidade*, Galaxia, 2005]

Era unha casa, se non de fartura, si de boa mantenza. [CORGA: T. Moure, *Herba moura*, Xerais, 2007]

Así pois, en XIADA *senón* figurará sempre ben como preposición ben como conxunción subordinante, e *senón que* só o fará como locución conxuntiva subordinante. Non obstante, son moitas as dúbidas que esta unidade nos suscita en relación coa diferenza entre os usos exclusivos, substituíbles en xeral por *máis ca* (*Non come senón carne*) –clasificados tradicionalmente na clase das conxuncións– e os exceptivos, conmutables por *excepto* e *salvo* (*Senón ti, todos o fixemos*) –delimitados na clase das preposicións–, polo que agardamos que nun futuro non moi afastado outros investigadores profunden nestas cuestións e se poida chegar a unha solución, se non plenamente satisfactoria, si, polo menos, axeitada ás múltiples circunstancias que envolven estas partículas.

4.4.15.3.3. Quitado/quitando, sacado/sacando

Polo resumo que vimos no comezo desta macroepígrafe semella que para a delimitación destas unidades se seguen tres vertentes na filoloxía galega:

1) A que manteñen Costa Casas et al. (1988) e Freixeiro (2000), quen non as contan entre os elementos prepositivos nin conxuntivos. Esta consideración correspóndese coa que se mantén tamén na lingüística portuguesa (Cunha e Cintra, 1984 ou Vilela, 1999).

2) A que mantén a Real Academia Galega e mais os autores que a seguen (Álvarez, Regueira e Monteagudo, 1986; RAG/ILG, 1995 e 2003; Hermida, 2004) ao trataren os elementos simples como preposicións; a maiores, a combinación con *que* trátase como locución de tipo adversativo en Álvarez, Regueira e Monteagudo (1986) e, finalmente, nas *Normas* (RAG/ILG, 1995 e 2003), ademais do anterior, tamén se consideran as combinacións con *se* e *que* como locucións de tipo condicional.

3) A que seguen Álvarez e Xove (2002) ao consideralos conxuncións.

No corpus etiquetado non temos máis ca dous casos que se poderían encadrar aquí:

Sacado deleitarse cun alustro de envexa, nada lle pode facer. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

A partir dese momento, xa empecei a ter choio seguido e non parei, sacando unha hora para xantar, ata que se me botou a noite enriba. [CORGAetq: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Porén, no CORGA si temos algún exemplo máis para poder verificar se nestas catro unidades simples se dan dúas das características fundamentais que definen os elementos prepositivos: a invariabilidade e a posibilidade de levar por termo calquera unidade.

Comezaremos coas formas participiais, *sacado* e *quitado*, e tentaremos ver se é posible atopar nos mesmos contextos formas con variación xenérica e numérica.

Para o primeiro deles só documentamos un candidato, inmovilizado en masculino singular que ten como termo un pronome tónico, cuxa explicación, quizais, antes ca busca en ser termo anómalo de frase preposicional pode explicarse nunha construción de participio absoluto con suxeito pronominal nominativo:

E así foi, o único home, sacado eu, que montara nela fora o mecánico para probala e amañarlle algunha cousa... [CORGA: P. Martín, *Dende a muralla*, Xerais, 1990]

Constátase tamén a variante *de sacado*, mais emprégaa só un autor; recollemos un exemplo de cada obra das que temos metidas no corpus:

A Noiteboa retirou das rúas a todo becho vivente, de sacado ós gatos. [CORGA: B. Iglesias Araúxo, *Vento de seda*, Xerais, 1992]

De sacado o Ourense anal, que poboa as noites das beirarrúas de Correos, ningún outro signo resta de pasados esplendores, das pillerías do café-cantante ou do espírito copulativo que moraba na rúa do Vilar, inspirando a recuas de peós de albanel e de cavadores que saltaban nas meretrices, sobre o abismo do mundo escindido, anulando a distancia de Eros. [CORGA: B. Iglesias Araúxo, *A historia escríbese de noite*, Galaxia, 2001]

Ninguén se relaciona de balde cun impedido (de sacado a nai, pero a vella era un apéndice de si propio, ela tamén se esnafrara contra a mediana), nin a amantísima Ida lle daría corte de non mediaren os escritos sobre a eutanasia que o empurraran ao presenteiro da celebridade. [CORGA: B. Iglesias Araúxo, *Contos da Baiuca*, Espiral Maior, 2006]

As ocorrencias con formas participiais de *quitar* son algo máis numerosas, ademais de documentarse as variantes flexivas en feminino e plurais:

Quitado o caso da carne porcina e, en menor medida, o leite, o grado da cobertura galega da expansión gandeira española foi sensiblemente superior á participación galega nas producións gandeiras españolas en 1962. [CORGA: X. Colino e E. Pérez Touriño, *Economía campesiña e capital. A evolución da agricultura galega (1960-1980)*, Galaxia, 1993]

Sabes que, quitado o de durmires enriba miña no medio da función e non chegares á fin, gustoume. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]

Todo o brañal, quitado o cacho de terra firme onde viven as mulleres, é unha inmensa trampa da que non se pode saír con vida se un se aventura a intentalo sen axuda. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

O seu monto foi substancioso, xa que nun ano sumaron, quitadas aínda as da carne e as dos panos, nada menos que 50.930 mrs(229). [CORGA: A. López Carreira, *Ourense no século XV. Economía e sociedade urbana na Baixa Idade Media*, Xerais, 1991]

Quitadas as experiencias de Armando Cotarelo Valledor nos anos vinte e dalgunha función esporádica nos anos sesenta e setenta, a nosa universidade xamais foi receptiva á arte teatral. [CORGA: *TN1997-05/37*]

Mais ela era boa muller, quitadas as súas faltas, como calquera. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]

Asegúroche que as miñas amizades, quitadas un par de amigas, son un fato de parvos. [CORGA: S. de Toro, *Calzados Lola (O dano e a absolución)*, Xerais, 1998]

Quitada a obrigatoriedade, o celibato podía recuperar o valor de protesta contra a deshumanización do amor e do sexo, contra as distintas formas de escravitude e cousificación da sexualidade. [CORGA: X. Alvilaes, *Memorial de agravios*, Galaxia, 1992]

Fóra destes ámbitos, no século XIX non é doado atopar exemplos de anuncios en galego, quitados aqueles que poñen en coñecemento do público algunha novidade editorial neste idioma. [CORGA: X. Cermeño, *A publicidade. A fábrica dos desexos*, Xunta de Galicia, 1993]

Na festa Nacional do 12 de Outubro, antes "Festa da Raza" estiveron presentes, e en posición de firmes, ademais do goberno central, os presidentes das autonomías, quitados tres: Pujol, Fraga e Ibarretxe. [CORGA: *ANT2002-10-17/5*]

Non obstante, fóra os casos nos que *quitado* concorda en xénero e número co sintagma que o segue (exemplos anteriores), onde si parece que estamos ante un participio, atopámonos tamén con secuencias nas que *quitado* aparece inmovilizado en masculino singular e non mantén concordancia co sintagma co que se relaciona:

E non ten antecedentes penais, nin sequera multas sen pagar, quitado esta. [CORGA: X. Borrazás, *Na maleta*, Sotelo Blanco, 2000]

O azucre, o sal e o deterxente fixeran pedras coa forma dos paquetes que os contiñan e as plantas foran morrendo unha atrás doutra, quitado os cactos. [CORGA: X. Borrazás, *Na maleta*, Sotelo Blanco, 2000]

E tiña o siso posuído por un desexo máis grande cá monotonía do traballo na empresa e das relacións cotiás que constituían a súa célere vida, quitado Marga e algunha que outra esmorgada pasaxeira que deixaba un pouso de amargor no espírito e o corpo baldado para varias xornadas. [CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

E novamente documéntase tamén a variante coa preposición *de* anteposta, mais xa non nun só autor, o que nos ofrece máis garantías:

De quitado que algúns ás veces lle chamaban Moisés e logo pedían perdón sorrindo polo equívoco, o Rubén non notaba nada raro no trato da xente, e seguiu facendo a súa vida de cotío. [CORGA: D. X. Cabana, *Vidas senlleiras*, Xerais, 1994]

Endémicamente alérxico ós cataláns (de quitado os socios do Español RCD) e non digamos ós vascos (agás os carlistas ou o adegueiro Olarra), a política de alianzas do PFP coas minorías nacionalistas nos próximos comicios basearase nos seus paisanos. [CORGA: *GII1996-01/34*]

De quitado Rajoy, poucas estrelas da futura vida parlamentaria se intúen no grupo de deputados e senadores galegos do PP. [CORGA: *GII1996-04/34*]

Se agora atendemos aos termos das supostas frases preposicionais encabezadas por *sacado* ou *quitado*, comprobamos que tanto pode ser un pronome tónico en caso nominativo (*eu*), como unha frase nominal (*os cactos*), unha substantiva (*esta*), outra

preposicional (*ós gatos*), ou mesmo unha cláusula de infinitivo (*deleitarse...*) ou unha cláusula con *que* ou *se*.

Como acabamos de comprobar, a lingua non é algo estático senón vivo, na que un elemento dado, clasificado orixinariamente como participio (*sacado/a(s)*, *quitado/a(s)*), vai modificando en ocasións o seu comportamento ata chegar a poder desempeñar unha función distinta e asumir, polo tanto, o rol doutra categoría morfolóxica. Na nosa opinión, pois, este proceso tivo/ten lugar cos participios *sacado* e *quitado*, e proba diso son os casos inmovilizados en masculino singular que non presentan concordancia cos sintagmas que o seguen e nos que non podemos falar de construcións de participio absoluto, senón que temos que tratalos xa como frases preposicionais cuxo director é a preposición *sacado* ou *quitado*.

No sistema XIADA as unidades *sacado* e *quitado*, coas súas variantes *de sacado* e *de quitado*, invariables en xénero e número, serán catalogadas na clase Preposición sempre que se perciba a deslexicalización e poidan ser conmutadas por *excepto* sen que varíe a significación da secuencia, o que as discrimina xa dos casos nos que esas unidades funcionan como verdadeiros participios. Dado que no exemplo que trañamos ao comezo non é posible a conmutación por outros participios de exclusión, coma *exceptuado* ou *prescindido de* con cambio de réxime, pero si por *excepto*, *menos* ou *salvo*, etiquetámolo como preposición cuxo lema é *sacado*:

Sacado deleitarse cun alustro de envexa, nada lle pode facer. [CORGAetq: A. Dobao, *Incertos*, Xerais, 2008]

Polo que respecta ás unidades con forma aparente de xerundios, *sacando* e *quitando*, non podemos acudir á flexión xenérica e numérica por ser unha forma non persoal na que non aplican estes atributos, mais, dado que se trata de formas de procedencia verbal, si podemos comprobar se os esquemas argumentais que as caracterizan como tales se seguen mantendo nas ocorrencias que, se seguimos os distintos manuais, se supoñen usos preposicionais²²⁴.

²²⁴ Nin no TILG nin no CTAG, os outros dous corpus do galego lematizados e etiquetados, aparece ningún caso para *sacando*, *quitando* na categoría Preposición. [Consultados o 30/03/2012]

Para iso, consultamos e extraemos do CORGA todas as ocorrencias de *quitando* e *sacando* que poidan ser conmutadas por *excepto* e converterse así en candidatas á análise como preposicións. Algúns exemplos son os seguintes:

Sacando a maquillaxe esaxerada e o pelo tinxido dun loiro chamativo, tiña enfronte unha auténtica beleza. [CORGA: Ú. Heinze, *Culpable de asasinato*, Xerais, 1993]

Lin libros, todos cantos había alí, libros que, sacando a borrallada, me permitiron moitas ideas sobre a sociedade, en concreto sobre o noso país, Galiza... [CORGA: M. Lourenzo González, *Arqueofaxia*, Xerais, 1995]

Naquela casa, sacando as catro sinxelas cadeiras da cociña, non había máis que dúas antigas butacas de coiro azulado. [CORGA: X. Fuentes Castro, *O pescador de cana*, Toxosoutos, 2006]

Digo a Deus, non a min nin a ningún outro prelado, quitando o Pontífice Romano, que todo o pode gobernar. [CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

Se o penso con detenemento miña nai era bastante máis forte que meu pai, el tiña, como é natural, forza física, pero quitando eso, miña nai era moito máis forte. [CORGA: C. Franco, *A lúa no cénit e outros textos*, Edicións do Castro, 1988]

Porque, quitando á miña nai, el era a última ponla que subsistía da parte inglesa da familia. [CORGA: M. A. Murado, *Memoria de derribos*, Galaxia, 1993]

"Quitando a Ablanedo, hoxe en día o Spórting non ten xogadores dese tipo e non sei como poderá saír da actual situación", asegura Manjarín que desexa que "eles non pasen ó final apuros, aínda que este domingo nós intentaremos mellora-lo xogo que vimos facendo e para que deste xeito poidamos ganar". [CORGA: *CG1995-02-11/103*]

Pero, quitando estes cambios lóxicos nas liñas que veñen de Madrid, haberá zonas que continúen esquecidas, como o interior da provincia de Lugo e Ferrolterra. [CORGA: *GH2007-03-06/1*]

E atopamos tamén a exclusión expresada co xerundio precedido da preposición *en*, malia que curiosamente só localizamos exemplos co verbo *quitar* e non con *sacar*:

Fernando Quintas: En quitando iso, era unha grandísima persoa. [CORGA: X. Flores, *Historia do Deportivo*, Xerais, 1993]

Lestón Fabeiro: Pero daquela case tódolos campos eran iguais, en quitando os dos grandes. [CORGA: X. Flores, *Historia do Deportivo*, Xerais, 1993]

Anque nós ben sabemos que de Praducelo non virá ninguén ao mítin... eiquí non hai máis que mixiriqueiros que beben adoito auga das pías das igrexas... / En quitando ao Perexe que,

amáis de capador dos bóns, é dos do bugallo amarguxado... [CORGA: X. F. López Arias, *A Negra / A caza das cascudas / As augas toldas / Terra sequeira*, Ediciós do Castro, 1979]

E, así mesmo, estes dous xerundios tamén expresan exclusión, aínda sen ser conmutables por *salvo* ou *excepto*, cando están acompañados de circunstanciais ou con pronomes enclíticos, feito inconcibible nunha preposición:

Era bon home, moi correcto, non sabía que facerlle... pero sacándoo do traballo sobre os blasóns familiares, non tiña máis temas de conversación. [CORGA: M^a X. Queizán, *A semellanza*, Sotelo Blanco, 1992]

Quitando a un lado estas comparacións que se me veñen á cabeza, o certo é que a de onte foi unha noite que non vou esquecer. [CORGA: X. López Rodríguez, *Días de desterro*, Xerais, 1996]

Se nos fixamos nos exemplos, tanto de *sacando* coma de *quitando*, observaremos que o esquema argumental é o mesmo: PREDICADO + COMPLEMENTO DIRECTO, sen que nos importe que a expresión deste último, se é animado, varía e pode aparecer con ou sen a preposición *a*.

Outro factor que nos afianza na crenza de que non estamos ante preposicións é a aparición doutros xerundios que tamén indican exclusión coma:

Os débitos a clientes, excluindo o efecto da cesión temporal de activos, presentan un crecemento interanual do 7,2 por cento. [CORGAetq: *CG1999-05-01/25*]

Neste senso, contrasta a austeridade dos mobles dos albergues co luxo que debeu presidir a decoración das sedes oficiais do Xacobeo, xa que a Sociedade a penas sobrepasou os 50 millóns en acondicionar as 19 residencias para peregrinos excluindo o Monte do Gozo, mentres que só a sala de xuntas e o mobiliario de alta dirección acaprou preto de 28 millóns, segundo as cifras que aparecen no informe do Consello de Contas. [CORGAetq: *TN1998-04/12*]

En liñas xerais, tódolos segmentos experimentaron aumentos nas matriculacións, exceptuando os vehículos comerciais lixeiros, que caeron un 1,83 por cento. [CORGAetq: *CG1996-02-10/98*]

Así pois, o sentido exceptivo que se documenta no xerundio de *sacar* e *quitar*, na nosa opinión, é deducible do valor fundamental que se lle concede a estes verbos: *Facer que [algo ou alguén] deixe de estar [co que estaba ou con quen estaba, sobre o*

que estaba, etc.] (DRAG 1997, 2012²²⁵ *quitar, sacar*, primeira acepción) e non se debe a unha recategorización en preposición. En conclusión, a non ser que nos xurda algunha secuencia con un destes xerundios seguidos só por pronome tónico recto (*eu, ti, el...*), en cuxo caso sería imposible pensar que estamos ante un Complemento Directo e non nos quedaría máis remedio que recoñecelos como membros da clase Preposición, no sistema de XIADA serán formas verbais xerundiais.

4.5. Conclusións

Entre as principais conclusións que podemos extraer da lectura deste capítulo destacan as seguintes:

1) Os textos do CORGA demostran que o proceso de normalización do galego, por desgraza, aínda non se acadou na súa totalidade. Isto é obvio sobre todo no nivel que máis traballamos en XIADA, o léxico, onde a presenza de voces alleas á norma segue a ser unha constante. Con todo, por sorte para o futuro da nosa lingua, tamén se aprecia un rexistro máis coidado a partir do 2004, patente na redución considerable deste tipo de vocábulos.

2) Teñen aínda bastante que mellorar os tradutores e correctores actuais para podermos prescindir dunha última lectura atenta, a modo de supervisión, antes de que calquera texto saia das nosas mans. O contrario non só vai en detrimento do bo facer de calquera profesional, senón que tamén constitúe unha agresión á lingua e, en última instancia, á conciencia lingüística do lector final.

3) Normativizar o corpus de análise, é dicir, corrixilo para adaptalo á normativa actual, así como prescindir da análise de elementos problemáticos tales como cifras ou datas, é unha solución artificial e innecesaria. Innecesaria, por canto a mera existencia do CORGA etiquetado demostra que é posible traballar con texto sen ter que normativizalo e tamén sen ter que pasar por riba de unidades tanxenciais nin reducilas a un formato único, como sucede habitualmente coas cifras e as datas para as que se crean clases *ad hoc*. Artificial, en caso de os textos seren manipulados, por canto supoñería non traballar con lingua real, así como porque tampouco permitiría coñecer o alcance real do etiquetador, ademais de limitar o desenvolvemento pleno das ferramentas informáticas que deben aplicárselles aos documentos. Así pois, nada empece a

²²⁵ Dispoñible en <http://www.realacademiagalega.org/diccionario> [Consultado o 10/04/2012]

aplicación do Etiquetador e lematizador do sistema XIADA en calquera outro tipo de textos reais actuais.

4) Os corpus textuais –incluso sen anotar pero, por suposto, moito máis útiles en relación co nivel de anotación que presenten– confirmanse como ferramentas óptimas para levar a cabo estudos de tipo léxico, morfosintáctico ou pragmático.

5) As moitas dúbidas que manifestamos na segunda parte deste capítulo en relación con distintos aspectos da análise realizada no CORGAetq e a súa oportuna xustificación constatan a escaseza de obras de consulta, sobre todo traballos específicos, ás que acudir para completar as lagoas que conteñen os traballos xerais. De ningún xeito é unha crítica a ningunha das gramáticas galegas do mercado; os seus autores merecen o recoñecemento de todos os que loitamos, dun xeito ou doutro, pola defensa da lingua galega. Mais é preciso seguir afondando; é necesario seguir traballando sobre aspectos xa descritos (por exemplo, as distintas categorías gramaticais ou os elementos discursivos) e describir polo miúdo outros aspectos só tratados superficialmente (as partículas exceptivas, criterios de delimitación entre adverbios seguidos de preposición e locucións prepositivas, a interpolación de elementos, a polaridade, etc., etc.), sen esquecer a coherencia e mellora substancial das obras básicas que guían a normativa actual: *Normas ortográficas e morfolóxicas do idioma galego* (2003), *Vocabulario Ortográfico da Lingua Galega* (2004), *Diccionario castelán-galego da Real Academia Galega* (2004) e *Diccionario da lingua galega* (2012), así como a publicación dunha gramática académica descritiva o máis completa posible e rigorosa.

Conclusións

A diversidade temática dos catro grandes apartados que constitúen este traballo (establecemento dun etiquetario, construción dun lexicón, características da etiquetaxe automática e descrición dos criterios xerais aplicados na etiquetaxe e lematización manuais) provoca que cada un deles inclúa unhas conclusións específicas, de aí que aquí se recolla unicamente unha valoración global e sintética do sistema XIADA.

A elaboración de recursos lingüísticos destinados ao procesamento da linguaxe natural, calquera que sexa a lingua da que se trate, obriga a ver esta dende unha perspectiva computacional. A etiquetaxe de corpus non é allea a este principio xeral. Para poder realizala cómpre conxugar os intereses lingüísticos coas necesidades computacionais.

A ferramenta de análise automática escollida precisa coñecer de antemán o etiquetario que debe empregar na súa análise, así como o código que rexe a orde dos elementos, a súa combinatoria, etc., sen esquecermos do coñecemento do léxico principal da lingua en cuestión. Os obxectivos lingüísticos determinan, en certo modo, a complexidade de cada un dos tres bloques anteriores.

O *Etiquetador/Lematizador do galego actual* (XIADA) nace para etiquetar os documentos que conforman o *Corpus de Referencia do Galego Actual* (CORGA) e permitir que se dea un salto cualitativo no seu sistema de consultas. O obxectivo xeral que se persegue é facilitar a recuperación de información a través dunha aplicación web mediante a consulta por *forma*, *lema*, *etiqueta morfosintáctica* ou calquera combinación destas; co engadido, naturalmente, das prestacións do CORGA: tipo de documento sobre o que interesa realizar a consulta (*medio*), período cronolóxico que se quere abranguer (*Dende... Ata*), selección temática (*área temática*, *subárea temática*), parte estrutural do documento no que realizar a busca (*corpo*, *titular*, *encabezamento*, *pé de foto...*), etc.

O desexo de satisfacer as necesidades dun amplo abano de usuarios potenciais, fornecéndoos do material preciso para desenvolveren as súas investigacións ou obterem

datos puntuais, está auspiciado pola orixe mesma do proxecto: a construción dun corpus que pretende ser de referencia dos usos lingüísticos do galego actual, promovido e financiado por unha entidade pública –a Xunta de Galicia a través da Secretaría Xeral de Política Lingüística, de quen depende o Centro Ramón Piñeiro para a Investigación en Humanidades–, gratuíto e de acceso libre, esixe que o produto resultante sexa accesible a todos, de calidade, sinxelo na expresión, pormenorizado no contido, graduable na obtención dos datos... Estas premisas motivan o deseño dun sistema ambicioso, reflectido esencialmente no etiquetario que se emprega. Case 400 etiquetas diferentes, en concreto 383, resultado de combinar as categorías gramaticais recoñecidas en XIADA cos seus atributos e valores pertinentes, dan conta, en teoría, da caracterización morfosintáctica de calquera palabra galega e, na práctica da etiquetaxe de corpus, permiten a adecuación desa caracterización a un contexto específico.

A ambición materializada no etiquetario de XIADA guíaa, así mesmo, o anhelado de dar conta da realidade lingüística presente nos documentos do CORGA, reflexo á súa vez da realidade lingüística galega actual por ser considerado este un corpus de referencia. Así é que non só se ambiciona a etiquetaxe cunha alta porcentaxe de acerto de textos en galego normativo actuais, senón tamén o recoñecemento e caracterización daqueles vocábulos que, por seren considerados castelanismos, populismos, hipergaleguismos, etc., a norma culta non acepta. Grazas a esta minuciosidade é posible, por exemplo, estudar o uso e distribución en galego das formas en *-ra* e *-se* no seu valor de pretérito de subxuntivo a partir dos datos extraídos do *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq). Ou o plural característico do bloque central para as formas rematadas en *-n*, a utilización da preposición *cara* no sitio da locución *cara a*, o emprego de formas pronominais de dativo en vez das de acusativo en construcións moi específicas, a diminución no emprego das segundas formas do artigo a raíz da última revisión da normativa oficial, etc., etc. Naturalmente, todo isto sen que se obstaculice de ningún xeito as consultas máis simples e, o que cremos tamén fundamental, sen ter que coñecer previamente o código empregado na etiquetaxe, dado que o etiquetario traducíuse a unha interface na que, mediante un menú sinxelo e comprensible, o usuario selecciona a categoría gramatical do(s) elemento(s) que lle interesa consultar e os atributos pertinentes para esa clase, determinando el mesmo o nivel de profundidade que precisa.

A exhaustividade na análise morfosintáctica non se logra só cun etiquetario escrupuloso, nin moito menos está relacionada co sistema que permite consultar os datos e a súa posterior extracción, senón que require de dúas fontes das que tomar os datos relativos á clasificación e caracterización das palabras que compoñen os textos: o lexicón e o corpus de adestramento.

Polo que se refire ao primeiro, o lexicón, conseguimos reducir a unha gramática formal a maior parte do acervo léxico do galego actual, non soamente atendendo aos vocábulos normativos, senón tamén integrando no sistema aqueles outros rexeitados ou non aconsellados pola normativa oficial vixente máis frecuentes no CORGA. A distinción entre lemas e raíces mais a creación dos grupos de derivación vistos no Capítulo 2 posibilita a concreción do etiquetario elixido, aplicado ao léxico galego, sen ter que introducir cada unha das formas gráficas. Así pois, a variación flexiva definitoria da nosa lingua redúcese formalmente en XIADA co obxectivo de facilitar a implementación do léxico. Mais o propósito do lexicón, malia ser o primordial, non é exclusivamente servir de dicionario do etiquetador. A información relativa a tipo de léxico, fonte onde se documenta e normatividade da forma e do lema propicia que o lexicón poida considerarse un recurso lingüístico con entidade propia, relevante para outras finalidades tales como base dun corrector ortográfico, conxugador verbal, detección de neoloxismos, produtividade de sufixos, etc.

Se reducir formalmente a flexión nos adxectivos e substantivos, así como a conxugación do paradigma verbal, nos brindou certa satisfacción, o auténtico logro do sistema XIADA, na nosa opinión, é a materialización da diferenza entre *unidade*, *constituínte*, *forma* e *lema*, que se orixina no lexicón, e a súa aplicación na etiquetaxe de corpus. É dicir, lograr identificar e clasificar previamente as unidades léxicas sen que a súa aparición nunha forma ortográfica simple, nunha amálgama ou nunha unidade multipalabra constituía un atranco insalvable para o seu recoñecemento. Tal como describimos no Capítulo 2, o galego presenta unha inxente cantidade de amálgamas entre elementos lingüísticos de distinto tipo, destacando entre estas, precisamente por ser as únicas que non constitúen unha serie pechada, a combinación de todas as formas verbais, menos o participio, con pronomes enclíticos e mais a combinación das formas rematadas en *-r* ou *-s* coas segundas formas do artigo. É dicir, formalmente, a conxugación verbal cando inclúe algún enclítico ou segunda forma do artigo difire da

conxugación verbal simple por canto se produce en ocasións a supresión de caracteres ou a variación na acentuación gráfica da raíz verbal ao ser considerados a efectos gráficos ambos os elementos unha única unidade. A reduplicación do paradigma verbal tendo en conta estes cambios na grafía das raíces verbais, xunto cunha relación das posibles combinacións de pronomes enclíticos e as etiquetas que lles corresponderían en cada caso, unido a un dicionario inverso creado con esas formas verbais e relacionado coa conxugación normal permite en XIADA que se identifique a unidade e se caractericen correctamente –salvo casos de ambigüidade nos que o sistema pode fallar– os constituíntes da amálgama resultante.

A especificación dos diferentes clíticos que concorren coa forma verbal, así como os posibles valores de cada un deles, incrementa o valor do lexicón e, sobre todo, do corpus desambiguado, dado que, fronte a outros corpus etiquetados en galego coma o CTAG ou o TILG por exemplo, o CORGAetq non só contén información concreta sobre a etiqueta que lles corresponde en función do contexto no que se localizan, senón que permite realizar consultas a través da aplicación en liña sobre unha parte ou a totalidade dun valor determinado. Naturalmente, a utilidade que un sistema deste tipo pode prestar en xeral no procesamento da linguaxe natural é evidente. Pénsese, por exemplo, na tradución automática.

O corpus de adestramento, pola súa parte, accesible ao público case na súa totalidade no *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq), é o resultado palpable da aplicación da etiquetaxe automática, revisada manualmente por un lingüista, na que se unen todos os recursos lingüísticos aquí descritos: etiquetario, lexicón, regras lingüísticas, combinacións de enclíticos, relación de contraccións, dicionario inverso de raíces verbais que se combinan con pronomes enclíticos...

Varios son os motivos de satisfacción que nos proporciona o corpus de adestramento. Por unha banda, o logro dun editor XML que nos permite mediante DTD estruturar e representar a información que desexamos para os distintos tipos de unidades existentes: simples, amalgamadas e multipalabra. Por outra, a recuperación a través dunha aplicación de consulta web pública da información que figura no corpus: unidade, forma, etiqueta e lema de ata catro unidades léxicas. Finalmente, o volume e exhaustividade na etiquetaxe do corpus: case medio millón de formas ortográficas na

versión 2.5, en concreto 499 202. Nada de todo isto sería posible nin visible sen o traballo inestimable do informático do proxecto, polo que os méritos deben recaer tamén nel.

A construción do corpus de adestramento foi –e é, posto que non se deu aínda por concluído– un desafío constante, xa que implica o tratamento e a caracterización de todas as unidades que poden aparecer nun texto, incluídas aquelas que non adoitan recollese nos manuais (siglas, abreviaturas, fórmulas, símbolos, acurtamentos...), as características dunha lingua en proceso de normalización (dialectalismos, hipergaleguismos, castelanismos...) e as típicas resultado de erros de diferente tipo (erratas, ultracorreccións...). Ás dificultades que de por si provoca a etiquetaxe morfosintáctica (diferenciación categorial entre elementos próximos, atribución de valores, identificación de unidades multipalabra, etc.) xúntase o requisito da coherencia, da sistematización na etiquetaxe, imprescindible para que o etiquetador estatístico presente unha taxa alta de acerto nos seus resultados. A verificación dunha taxa de acerto do 96% é gratificante, pero pensamos que tamén mellorable. O camiño a seguir para poder chegar ao teito de acerto dun etiquetador estatístico coma o de XIADA debe dirixirse, polo tanto, cara a rematar o corpus de adestramento e elaborar regras lingüísticas que corrixan erros sistemáticos de etiquetaxe e axuden o etiquetador a decidir nos casos de ambigüidades segmentais complexas, sen esquecer o desenvolvemento dun módulo máis avanzado para a detección e caracterización de nomes propios.

Finalmente, tamén é xusto recoñecer que a pesar das satisfaccións que o sistema nos ofrece e dos estímulos que nos proporciona porque sempre hai algunha unidade, algunha combinación, algunha cuestión lingüística en definitiva que suscita o interese por saber máis dela, a desambiguación manual do corpus provocou –e provoca– moitas dúbidas, creando inseguridade sobre algunhas das decisións tomadas; non obstante, cremos que aínda así se trata dun recurso moi útil para o estudo da lingua galega, que ademais axuda a situala entre aquelas que gozan de recursos computacionais, favorecendo a incorporación do galego ao mundo das novas tecnoloxías da información.

APÉNDICES

Apéndice A

Etiquetario de XIADA

		c- coordinante s- subordinante 0- non aplica	m-masculino f-feminino a-masculino/feminino n-neutro 0- non aplica	s-singular p- plural a- singular/plural 0- non aplica	c-comparativo s-superlativo 0-non aplica	1-primeira 2-segunda 3-terceira a-primeira/terceira 0- non aplica	a-acusativo d-dativo n-nominativo p-preposicional f- formante léxico o-outros casos	p-presente i-copretérito l-antepretérito f-futuro e-pretérito c-pospretérito 0-non aplica	i-indicativo s-subxuntivo m-imperativo f-infinitivo x-xerundio p-participio	s-singular p-plural a-singular/plural	n-non determinante d-determinante a-determinante/non determinante
Categoría	Tipo	Subtipo	Xénero	Número	Grao	Persoa	Caso	Tempo Verbal	Modo	Posuidor	Valor
Substantivo (S)	c-común p-propio		m, f, a, 0	s, p, a, 0							
1	2		3	4							
Adxectivo (A)			m, f, a	s, p, a	c, s, 0						
1			3	4	2						
Verbo (V)			m, f, 0	s, p, 0		1ª, 2ª, 3ª, a, 0		p, i, l, f, e, c, 0	i, s, m, f, x, p		
1			5	6		4		2	3		
Preposición (P)											
1											
Conxunción (C)	c-coordinante s-subordinante										
1	2										
Adverbio (W)	n-nuclear m-modificador a-nuclear/modificador r-relativo g-interrogativo-exclamativo										
1	2										
Artigo (D)	d-determinado i-indeterminado		m, f	s, p							
1	2		3	4							
Demostrativo (E)			m, f, n	s, p							d, n
1			3	4							2
Relativo (T)			m, f, a	s, p, a							d, n
1			3	4							2
Posesivo (M)			m, f	s, p		1ª, 2ª, 3ª				s, p, a	d, n
1			5	6		3				4	2
Indefinido (I)			m, f, a, 0	s, p, a, 0							d, n, a
1			3	4							2
Numeral (N)	c-cardinal o-ordinal		m, f, a	s, p							d, n
1	2		4	5							3
Pronome (R)	t-tónico a-átono		m, f, n, a	s, p, a		1ª, 2ª, 3ª	a, d, n, p, f, o				
1	2		5	6		4	3				
Interrogativo-Exclamativo (G)			m, f, a	s, p, a							d, n
1			3	4							2
Locución (L)	a-adverbial p-prepositiva c-conxuntiva	c, s, 0									
1	2	3									
Interxección (Y)											
1											
Signo de puntuación (Q)											
1											
Categoría Periférica (Z)	a-abreviatura f-fórmula g-sigla s-símbolo o-outros tipos		m, f, a, 0	s, p, a, 0							
1	2		3	4							

Apéndice B

Listaxe de etiquetas empregadas no CORGAetq

A listaxe que presentamos está ordenada alfabeticamente pola etiqueta, a continuación da cal se desenvolve o seu valor e se engade un ou varios exemplos extraídos do propio *Corpus de Referencia do Galego Actual etiquetado* (CORGAetq versión 2.1, que consta de 386.498 formas ortográficas ou se se prefire 426.264 elementos gramaticais).

A presente relación consta de 299 etiquetas distintas, as cales resultan de combinar as categorías recoñecidas no etiquetario de XIADA cos seus atributos e valores correspondentes, aplicados á anotación morfosintáctica do CORGAetq. As etiquetas aquí recollidas só representan o 78,07% do total de etiquetas que prevemos utilizar na anotación de corpus en galego (383).

ADJECTIVO (A) (16 etiquetas)

A0aa→Adxectivo, grao non aplica, masculino/feminino, singular/plural: ...*a demanda mundial de off-shore centrarase nos próximos anos neste último tipo de plataformas.*

A0ap→Adxectivo, grao non aplica, masculino/feminino, plural: *Somos diferentes e universais e a nosa forma de ver a vida...*

A0as→Adxectivo, grao non aplica, masculino/feminino, singular: *É moi fráxil neste senso.*

A0fp→Adxectivo, grao non aplica, feminino, plural: *As continuas noticias sobre naufraxios ...*

A0fs→Adxectivo, grao non aplica, feminino, singular: *A peseta queda como unha lembranza bonita e infantil...*

A0mp→Adxectivo, grao non aplica, masculino, plural: *...nos conflitos bélicos, aos pequenos países...*

A0ms→Adxectivo, grao non aplica, masculino, singular: *nacionalismo galego moderno...*

Acas→Adxectivo, comparativo, masculino/feminino, singular: *o número e gravidade de conflitos bélicos e atentados terroristas foi menor neste ano...*

Acfp→Adxectivo, comparativo, feminino, plural: *...gozando de maiores oportunidades...*

Acfs→Adxectivo, comparativo, feminino, singular: *...maior flexibilidade e axilidade...*

Acmp→Adxectivo, comparativo, masculino, plural: *...proclamaba un dos mellores eslogans do nacionalismo...*

Acms→Adxectivo, comparativo, masculino, singular: *O sábado foi o peor día con sete accidentes e sete mortos.*

Asfp→Adxectivo, superlativo, feminino, plural: *...as obras son máquinas complexísimas.*

Asfs→Adxectivo, superlativo, feminino, singular: *A cooperativa debe preocuparse de marca-la mínima marxe económica...*

Asmp→Adxectivo, superlativo, masculino, plural: *Os prezos máximos das gasolinas baixarán a partir de mañá...*

Asms→Adxectivo, superlativo, masculino, singular: *¡Carísimo e cunha calor que fundía o queixo!*

CONXUNCIÓN (C) (2 etiquetas)

Cc→Conxunción coordinante: *tres de cada dez libros que se producen quedan sen vender, ben nos almacéns das editoriais, ben nos estantes das librerías.*

Cs→Conxunción subordinante: *Damián Villalaín, responsable de Galaxia, considera que, unque a Xunta e as editoriais galegas colocaron "unha boa mostra" en Fráncfurt, cómpre levar a cabo esa promoción de autores.*

ARTIGO (D) (8 etiquetas)

Ddfp→Artigo, definido, feminino, plural: *As familias gastan menos ca en 1990*

Ddfs→Artigo, definido, feminino, singular: *Parámetros que permiten definir a saúde*

Ddmp→Artigo, definido, masculino, plural: *A reforma do IRPF supuxo un aforro de 800.000 millóns para os contribuíntes*

Ddms→Artigo, definido, masculino, singular: *A iso haberá que engadir o impacto na capacidade emisora e receptora...*

Difp→Artigo, indefinido, feminino, plural: *No 1994 o banco presentaba unhas perdas de 371 millóns de pesetas...*

Difs→Artigo, indefinido, feminino, singular: *Referiuse a que se trata dunha plataforma ideal para os interesados...*

Dimp→Artigo, indefinido, masculino, plural: *...hai uns proxectos de desenvolvemento que se manteñen", explicou Rafael Miranda.*

Dims→Artigo, indefinido, masculino, singular: *No futuro, alguén lembrará que un día tivemos tigres de ouro...*

DEMOSTRATIVO (E) (9 etiquetas)

Edfp→Demostrativo, determinante, feminino, plural: *Estas medidas de tipo microeconómico son das poucas dispoñíbeis...*

Edfs→Demostrativo, determinante, feminino, singular: *Essa idea parte do descoñecemento.*

Edmp→Demostrativo, determinante, masculino, plural: *E se en Karaikal corre algo estes días é o alcohol.*

Edms→Demostrativo, determinante, masculino, singular: *Naquel momento e, desde logo, neste momento, non teño ningunha intención de cambiar.*

Enfp→Demostrativo, non determinante, feminino, plural: *A todas estas cabería engadir aquelas valoracións que consideran...*

Enfs→Demostrativo, non determinante, feminino, singular: *Asegura que esta é unha idea que "se leva anos acariñando"...*

Enmp→Demostrativo, non determinante, masculino, plural: *Ambos os candidatos apelan aos votantes para que estes sexan...*

Enms→Demostrativo, non determinante, masculino, singular: *Ese foi un dos temas que máis discutimos na Asamblea Constituínte.*

Enns→Demostrativo, non determinante, neutro singular: *Isto coloca a Galicia na cuarta comunidade con menos investimento...*

INTERROGATIVO-EXCLAMATIVO (G) (11 etiquetas)

Gdaa→Interrogativo-exclamativo, determinante, masculino/feminino, singular/plural: *"Qué máis quixesemos nosoutros", sinalou o presidente de Erimsa.*

Gdfp→Interrogativo-exclamativo, determinante, feminino, plural: *Rato explicou onte que o Goberno "está estudando qué decisións debe tomar..."*

Gdfs→Interrogativo-exclamativo, determinante, feminino, singular: *E, que casualidade!*

Gdmp→Interrogativo-exclamativo, determinante, masculino, plural: *Que criterios segue para escoller cada programa?*

Gdms→Interrogativo-exclamativo, determinante, masculino, singular: *Con que obxectivo nace o Sindicato de Xornalistas de Galicia?*

Gnaa→Interrogativo-exclamativo, non determinante, masculino/feminino, singular/plural: *Que debemos esperar del?*

Gnas→Interrogativo-exclamativo, non determinante, masculino/feminino, singular: *Quen vai marcar agora a política no concello?*

Gnfp→Interrogativo-exclamativo, non determinante, feminino, plural: *Cales son as principais dificultades técnicas da guitarra?*

Gnfs→Interrogativo-exclamativo, non determinante, feminino, singular: *Cal considera a mellor escola para formar guitarristas?*

Gnmp→Interrogativo-exclamativo, non determinante, masculino, plural: *Cáles son estes factores:*

Gnms→Interrogativo-exclamativo, non determinante, masculino, singular: *¿Cál é o papel das APA?*

Categoría-> INDEFINIDO (I) (17 etiquetas)

Iafp→Indefinido, determinante/non determinante, feminino, plural: *...baixa entre 3 e 4 pesetas as gasolinas e outras dúas o gasóleo*

Iafs→Indefinido, determinante/non determinante, feminino, singular: *Boa como calquera outra.*

Iamp→Indefinido, determinante/non determinante, masculino, plural: *...engloba este aspecto e moitos outros e nós pretendemos...*

Iams→Indefinido, determinante/non determinante, masculino, singular: *...controla o 33,3 por cento da empresa e outro tanto está en poder da empresa norueguesa NorKvarts.*

Idap→Indefinido, determinante, masculino/feminino, plural: *indicou que nove de cada dez asistentes a este evento...*

Idas→Indefinido, determinante, masculino/feminino, singular: *...creou a figura do "ciber-delegado" para que calquera internauta , ademais de seguir en directo o congreso, puidera enviarlle mensaxes á dirección no transcurso dos debates.*

Idfp→Indefinido, determinante, feminino, plural: *Despois de varias probas realizadas en Madrid...*

Idfs→Indefinido, determinante, feminino, singular: *Sen embargo, a adxudicación de tal cantidade fala ben claro da viabilidade...*

Idmp→Indefinido, determinante, masculino, plural: *O 11-S acentuou o proceso de globalización en todos os ámbitos*

Idms→Indefinido, determinante, masculino, singular: *Esta cifra supón tan só o 8% do total investido no mesmo período...*

In00→Indefinido, non determinante, xénero non aplica, número non aplica: *Calquera persoa sabe que eses cartos non chegan para nada, soamente se trata de prórrogas de estudos infinitas", sentencia Ceferino Díaz.*

Inaa→Indefinido, non determinante, masculino/feminino, singular/plural: *Si, aínda que se trata dun festival non tan masivo a nivel de actuacións como hai outros, neste caso hai cinco ou seis grupos por día, que xa é bastante, pero que non é un número tan grande como para que o público vaia esgotado nin para que os grupos teñamos cen problemas coas horas de proba, os retrasos e demais.*

Inas→Indefinido, non determinante, masculino/feminino, singular: *Ninguén quere volver preto do mar.*

Infp→Indefinido, non determinante, feminino, plural: *Aínda que moitas das nosas propostas, como dixemos, foron aceptadas...*

Infs→Indefinido, non determinante, feminino, singular: *Galicia logra alcanzar unha das taxas de inflación máis baixas dos últimos anos.*

Inmp→Indefinido, non determinante, masculino, plural: *O problema é que moitos dos investimentos que se proxectan teñen...*

Inms→Indefinido, non determinante, masculino, singular: *Por outro, ACS/Dragados.*

LOCUCIÓN (L) (4 etiquetas)

La0→Locución, adverbial, tipo non aplica: *De feito, o 37º Congreso parecía estruturado por e para a mocidade militante.*

Lcc→Locución, conxuntiva, coordinante: *Mentres, os prezos subiron unha décima nas comunidades de Andalucía, Asturias, Estremadura, País Basco e mais Ceuta e Melilla.*

Lcs→Locución, conxuntiva, subordinante: *O Hospital da cidade de Santiago era a institución con maior presenza no Banco, xa que participaba con 294 accións.*

Lp0→Locución, preposicional, tipo non aplica: *Verbo da memoria do CES, Suárez-Llanos rexeitou emitir calquera avaliación acerca do seu contido.*

POSESIVO (M) (38 etiquetas)

Md1pfp→Posesivo, determinante, 1ª, plural, feminino, plural: *Aínda que moitas das nosas propostas...*

Md1pfs→Posesivo, determinante, 1ª, plural, feminino, singular: *...a nosa inequívoca vontade de colaborar...*

Md1pmp→Posesivo, determinante, 1ª, plural, masculino, plural: *As continuas noticias sobre naufraxios que sacoden os nosos mariñeiros impulsaron definitivamente esta iniciativa...*

Md1pms→Posesivo, determinante, 1ª, plural, masculino, singular: *Incluso o viño branco, entre o que destaca o noso Albariño...*

Md1sfp→Posesivo, determinante, 1ª, singular, feminino, plural: *...é o que pon nas miñas notas e o acordado nas actas"...*

Md1sfs→Posesivo, determinante, 1ª, singular, feminino, singular: *A miña postura sobre iso vén de longo...*

Md1smp→Posesivo, determinante, 1ª, singular, masculino, plural: *Nos meus tempos no Partido Nacionalista Galego...*

Md1sms→Posesivo, determinante, 1ª, singular, masculino, singular: *"O meu ofrecemento é o de xestionar este proxecto..."*

Md2pfs→Posesivo, determinante, 2ª, plural, feminino, singular: *Ante a vosq mirada expuxen...*

Md2pmp→Posesivo, determinante, 2ª, plural, masculino, plural: *Os vosos soldados están no noso poder.*

Md2pms→Posesivo, determinante, 2ª, plural, masculino, singular: *A sucursal aparecerá daquela no vos pensamento...*

Md2sfs→Posesivo, determinante, 2ª, singular, feminino, singular: *...pasar de recoller a colleita a ser dono da túa tenda...*

Md2smp→Posesivo, determinante, 2ª, singular, masculino, plural: *"Cando medras nunha lingua na que non aprendes a expresar os teus sentimentos, o mellor é aprender unha nova para facelo".*

Md2sms→Posesivo, determinante, 2ª, singular, masculino, singular: *...informacións respecto ao teu estado de saúde.*

Md3afp→Posesivo, determinante, 3ª, singular/plural, feminino, plural: *Como eran as súas clases?*

Md3afs→Posesivo, determinante, 3ª, singular/plural, feminino, singular: *Algúns entes inscribíronse trala súa concesión*

Md3amp→Posesivo, determinante, 3ª, singular/plural, masculino, plural: *Segundo os seus cálculos, Brasil terá que conseguir 63 mil millóns de dólares en 2003 para evitar un cesamento de pagos.*

Md3ams→Posesivo, determinante, 3ª, singular/plural, masculino, singular: *O seu volume de negocio aumentou arredor do 15%*

Md3pfp→Posesivo, determinante, 3ª, plural, feminino, plural: *Ademais, os grandes buques teñen xa compartimentadas as súas adegas para poder pecharse e amañarse en alta mar en caso de fuga", sinala o xerente de Kafloat.*

Md3pfs→Posesivo, determinante, 3ª, plural, feminino, singular: *Terán que ver cal é a súa responsabilidade en levar a crispación...*

Md3pmp→Posesivo, determinante, 3ª, plural, masculino, plural: *A empresa tamén mercou coches por valor de 52 millóns de pesetas, e dotounos dos seus correspondentes teléfonos...*

Md3pms→Posesivo, determinante, 3ª, plural, masculino, singular: *E no borrador do seu informe tampouco incluíron referencia ningunha á contía dos cartos que percibiu.*

Md3sfp→Posesivo, determinante, 3ª, singular, feminino, plural: *...fixera desta unha das súas principais actividades nos anos noventa.*

Md3sfs→Posesivo, determinante, 3ª, singular, feminino, singular: *...expresou a súa decepción pola falta de colaboración de EEUU.*

Md3smp→Posesivo, determinante, 3ª, singular, masculino, plural: *...plasmaron a súa visión particular do noso medio natural e os seus problemas.*

Md3sms→Posesivo, determinante, 3ª, singular, masculino, singular: *A viúva do militar, Susan, manifestou o seu alivio polo ditame...*

Mn1pfs→Posesivo, non determinante, 1ª, plural, feminino, singular: *Era cando soaba aquilo de "a terra é nosa e non de Fenosa".*

Mn1pms→Posesivo, non determinante, 1ª, plural, masculino, singular: *...nun país civilizado e moderno coma o noso".*

Mn1sfp→Posesivo, non determinante, 1ª, singular, feminino, plural: *...son tomas de posición miñas e por iso inclúoos...*

Mn1sfs→Posesivo, non determinante, 1ª, singular, feminino, singular: *A miña é unha posición de autoesixencia...*

Mn1smp→Posesivo, non determinante, 1ª, singular, masculino, plural: *...navegando libre polo deficiente organismo dalgún deses tristes devanceiros meus, imaxínoos tristes...*

Mn1sms→Posesivo, non determinante, 1ª, singular, masculino, singular: *...fai eco da vella reivindicación de Virginia Woolf de "un cuarto de seu" ampliándoa a "unha casa de meu".*

Mn3pfs→Posesivo, non determinante, 3ª, plural, feminino, singular: *...os hipermercados ampliaron a súa desde o 15 por cento de vendas en 1987 ta o 30,7 por cento que alcanzaron en 1994.*

Mn3pms→Posesivo, non determinante, 3ª, plural, masculino, singular: *O apoio estrutural e financeiro desas entidades culturais con prestixio de seu en Madrid son as que realmente sosteñen e lle dan febra ao festival.*

Mn3sfp→Posesivo, non determinante, 3ª, singular, feminino, plural: *...Ramón Piñeiro explicaba as súas sobre Galiza e o galeguismo...*

Mn3sfs→Posesivo, non determinante, 3ª, singular, feminino, singular: *...as únicas declaracións que fai a todo o séquito de acólitos que leva detrás súa coma avésporas chuchonas detrás dun bote de marmelada...*

Mn3smp→Posesivo, non determinante, 3ª, singular, masculino, plural: *...el tivo que buscar a vida nas Canarias, na cancha e non na construción como moitos paisanos seus, pero no fondo era tamén un emigrante...*

Mn3sms→Posesivo, non determinante, 3ª, singular, masculino, singular: *O histórico empresario asegura que está afastado do mundo da construción e que o seu é ocuparse da fundación...*

NUMERAL (N) (18 etiquetas)

Ncdap→Numeral, cardinal, determinante, masculino/feminino, plural: *Cada ano acaban a carreira en Galicia arredor de cen xornalistas.*

Ncdfp→Numeral, cardinal, determinante, feminino, plural: *Un estadio de cinco estrelas como figuraba na propaganda...*

Ncdfs→Numeral, cardinal, determinante, feminino, singular: *A extensión media das parcelas privadas é de unha ou dúas hectáreas...*

- Ncdmp→Numeral, cardinal, determinante, masculino, plural: *Esta cifra supón unha media de 283 euros por habitante...*
- Ncdms→Numeral, cardinal, determinante, masculino, singular: *Un ano despois do atentado contra as Torres Xemelgas...*
- Ncnap→Numeral, cardinal, non determinante, masculino/feminino, plural: *Hai vinte e cinco anos tiña 13.000 máis, e era a terceira de Galiza.*
- Ncnfp→Numeral, cardinal, non determinante, feminino, plural: *...ten como destino soamente seis das 2.860 empresas madeiras que operan en Galicia.*
- Ncnfs→Numeral, cardinal, non determinante, feminino, singular: *Na operación inclúese unha oficina do Urquijo en León, tres oficinas en Xixón, unha en Avilés e unha en Palencia...*
- Ncnmp→Numeral, cardinal, non determinante, masculino, plural: *Esta cifra supón unha media de 283 euros por habitante, fronte os 392 do conxunto do Estado.*
- Ncnms→Numeral, cardinal, non determinante, masculino, singular: *O mecanismo comercializarase no 2003*
- Nodfp→Numeral, ordinal, determinante, feminino, plural: *...as primeiras eleccións democráticas dende 1936...*
- Nodfs→Numeral, ordinal, determinante, feminino, singular: *O español é a primeira lingua.*
- Nodmp→Numeral, ordinal, determinante, masculino, plural: *"Os primeiros resultados foron bos, pero despois tivemos problemas...*
- Nodms→Numeral, ordinal, determinante, masculino, singular: *Os datos de crecemento do segundo trimestre...*
- Nonfp→Numeral, ordinal, non determinante, feminino, plural: *As reses máis abundantes son as primeiras, pero as segundas, de raza cachena, con cornos grandes e cor entre marelo e negro son, sen dúbida, moito máis fermosas.*
- Nonfs→Numeral, ordinal, non determinante, feminino, singular: *A primeira sería a necesidade das entidades de reducir os custos...*
- Nonmp→Numeral, ordinal, non determinante, masculino, plural: *No segundo caso, se quere que España estea entre os primeiros da moeda única o PP debe explicar" cómo vai logra-las reduccions do déficit,...*
- Nonms→Numeral, ordinal, non determinante, masculino, singular: *Rosalía de Castro e Murguía viviron no andar segundo desta casa que fora de alugueiro, como outras que tiveron na Coruña...*

PREPOSICIÓN (P) (1 etiqueta)

P→Preposición: *En Galicia o número de obxectores que responden á enquisa está entre 200 e 300, pero sabemos que son moitos máis.*

SINAL DE PUNTUACIÓN (Q) (18 etiquetas)

Q¡→Signo de puntuación ¡→¡*Meu Deus!*

Q!→Signo de puntuación !

Q¿→Signo de puntuación ¿:→¿*Como é posíbel acadar unha rendibilidade tan alta?*

Q?→Signo de puntuación ?

Q(→Signo de puntuación (→*A compra mínima sitúase nun lote de 40 árbores, a un prezo de 270 euros cada un (45.000 pesetas).*

Q)→Signo de puntuación)

Q"→Signo de puntuación "→*Con isto, aseguran en Agrobyte, "as árbores representan máis diñeiro que o propio terreo antes da plantación".*

Q'→Signo de puntuación '→'*ArtNatura': reivindicación estética da protección do medio de Galicia*

Q,→Signo de puntuación ,→*Está claro que non é o mesmo unha librería chapada en castaño, que un rolo de papel hixiénico.*

Q→Signo de puntuación →*O autor da versión foi Antón Valcárcel y López de Lemos (Ourense 1887-Madrid 1963), membro das Irmandades da Fala, que acabou sendo secretario do Consello Superior de Cámaras de Comercio.*

Q.→Signo de puntuación .→*Podía ser lóxico que se aconsellasen as reestructuracións.*

Q...→Signo de puntuación ...→*Museo, "institución permanente, sen espírito de lucro, que acopia, conserva, documenta, estuda, ensina de xeito didáctico...." que abre no verán do 2002.*

Q/→Signo de puntuación /→*A Éxito das mensaxes propagandísticas que fan creer á sociedade que paga máis que antes e/ou que a iniciativa privada ha de superar á provisión de servizos públicos.*

Q:→Signo de puntuación :→*A idea favorábel á redución de impostos pode deberse a varias causas:*

Q;→Signo de puntuación ;→*En todo caso, autonomía financeira é sinónimo de responsabilidade;*

Q[→Signo de puntuación [→*Ante iso, [Xosé Manuel] Seoane García [director de ENCE] anunciou que esperan novos ventos.*

Q]→Signo de puntuación]

Q_→Signo de puntuación _→*Por moita retórica que lle poñan, non é fácil ocultar que desde mediados dos 70, agás no bienio 86-88, Galicia sofre unha continuada caída do emprego _141 mil nos últimos sete anos_.*

Categoría-> PRONOME (R) (60 etiquetas)

Raa1ap→Pronome, átono, acusativo, 1ª, masculino/feminino, plural: *A creación da Dirección Xeral de Cooperación Exterior fainos ser optimistas, pero temos que recordar que o esforzo financeiro, ata o momento, ten sido escaso.*

Raa1as→Pronome, átono, acusativo, 1ª, masculino/feminino, singular: *En Xapón metinme nunha escola de arte floral moderna.*

Raa1fp→Pronome, átono, acusativo, 1ª, feminino, plural: *O cambio é revolucionario pero pensa que aínda queda moito por camiñar polo que "non podemos sentirnos satisfeitas nin quedar na autocompracencia".*

Raa1fs→Pronome, átono, acusativo, 1ª, feminino, singular: *Nunca me magoaría, é un pai marabilloso", dixo a moza ucraína.*

Raa1mp→Pronome, átono, acusativo, 1ª, masculino, plural: *Mais han de voar xa para que non nos pillen adormecidos.*

Raa1ms→Pronome, átono, acusativo, 1ª, masculino, singular: *síntome aliviado coma un cadelo ao que acaban de quitarlle...*

Raa2ap: Pronome, átono, acusativo, 2ª, masculino/feminino, plural: *"Encargado pola miña calidade de director de convocarvos e, sequera inmerecidamente, teño a honra, que reputo sinaladísima, de saudarvos no nome deste Consello de Administración, coma no meu propio".*

Raa2as→Pronome, átono, acusativo, 2ª, masculino/feminino, singular: *Cando buscas o que te une e non o que te separa, chegas a acordos positivos.*

Raa2ms→Pronome, átono, acusativo, 2ª, masculino, singular: *...pasar de recoller a colleita a ser dono da túa tenda, ou de limpar o chan das oficinas a converterte te en xefe".*

Raa3as²²⁶: Pronome, átono, acusativo, 3ª, masculino/feminino, singular: *o axente preguntoulle se posuía carné de conducir, e ó responderlle afirmativamente, advertiulle de que ía perder tres puntos.*

Raa3fp→Pronome, átono, acusativo, 3ª, feminino, plural: *Produciraas se a sociedade llo permite.*

Raa3fs→Pronome, átono, acusativo, 3ª, feminino, singular: *A harmonía compréndese mellor se a tocas.*

Raa3mp→Pronome, átono, acusativo, 3ª, masculino, plural: *Estes cartos non se outorgan a fondo perdido, senón que hai que devolvelos ao cuarto ano da súa concesión.*

Raa3ms→Pronome, átono, acusativo, 3ª, masculino, singular: *Cal dos seus profesores o marcou máis?*

Rad1ap: Pronome, átono, dativo, 1ª, masculino/feminino, plural: *Esas expresións que moi axiña nos parece imposíbel que non tivesen existido antes.*

²²⁶ Esta etiqueta desaparece do noso sistema e polo tanto tamén o fará en versións posteriores do CORGA etiquetado polos motivos que explicamos en 4.3.2.5. *Vacilacións ~ variacións no caso acusativo e dativo.*

- Rad1as→Pronome, átono, dativo, 1ª, masculino/feminino, singular: *Contábame Arturo Casas que viu un filme no que os protagonistas están de viaxe, estacionan, entran nunha taberna.*
- Rad1fs→Pronome, átono, dativo, 1ª, feminino, singular: *"Un día deixa de baixarme o período", sinalou e, despois de ir ó médico e de poñerse inxeccións, descubriron que estaba embarazada.*
- Rad1mp→Pronome, átono, dativo, 1ª, masculino, plural: *...as discográficas non invisten nas novas propostas e pídennos aos grupos que afrontemos todos os gastos para que eles só teñan que ocuparse da distribución".*
- Rad1ms→Pronome, átono, dativo, 1ª, masculino, singular: *"¿Como me van vir aquí dicir que van pechar unha fábrica e poñerme unha boleira?", preguntou o alcalde...*
- Rad2ap→Pronome, átono, dativo, 2ª, masculino/feminino, plural: *Eu non militei en ningún partido pero non teño incomenente en explicarvos o meu pensamento ideolóxico.*
- Rad2as→Pronome, átono, dativo, 2ª, masculino/feminino, singular: *...un traballo discográfico sérveche de excusa para dar concertos...*
- Rad3ap→Pronome, átono, dativo, 3ª, masculino/feminino, plural: *Entre estas eivas está a da comercialización, que lles impide superar o mercado local.*
- Rad3as→Pronome, átono, dativo, 3ª, masculino/feminino, singular: *Repetiu sen cesar a palabra 'gracias', dedicou e dedicáronlle numerosos eloxios, deulle a benvida a quen rexerá...*
- Rad3fp→Pronome, átono, dativo, 3ª, feminino, plural: *As árbores tamén empezan a recuperar o manto verde que lles arrebataron o vento e a friaxe do inverno.*
- Rad3fs→Pronome, átono, dativo, 3ª, feminino, singular: *En realidade, os médicos vironlle dous DIU, un ben posto, que ela mesma se colocou, "e outro no corpo".*
- Rad3mp→Pronome, átono, dativo, 3ª, masculino, plural: *...teñen a posibilidade de predicar a súa ignorancia e transmitirles aos demais a súa incultura.*
- Rad3ms→Pronome, átono, dativo, 3ª, masculino, singular: *... corresponderalle ao Goberno galego determinar os nomes de lugar.*
- Raf1ap→Pronome, átono, formante léxico, 1ª, masculino/feminino, plural: *Na miña primeira achega a esta Comisión de Investigación quixera dicirlles que esta catástrofe non desaparece porque nos empeñemos en mirar para outro lado.*
- Raf1as→Pronome, átono, formante léxico, 1ª, masculino/feminino, singular: *Se me buscan, aquí estou (se non me atrevo a desconectar ou quedar fóra de cobertura).*
- Raf1mp→Pronome, átono, formante léxico, 1ª, masculino, plural: *Isto fai que teñamos que tocar mellor, xa que sabendo o nivel que hai, os concertistas vémonos obrigados a acadalo.*
- Raf1ms→Pronome, átono, formante léxico, 1ª, masculino, singular: *...de repente convértome en EEUU en autor de Bertelsmann sen que ninguén desa empresa me escribise xamais unha carta.*

- Raf2as→Pronome, átono, formante léxico, 2ª, masculino/feminino, singular: *Por exemplo, se ti vas a un hotel facilmente te atoparás cun cadro moderno, non figurativo... é algo normal.*
- Rtn1as→Pronome, tónico, nominativo, 1ª, masculino/feminino, singular: *Galbraith di moitas cousas que eu non comparto.*
- Rtn1ap→Pronome, tónico, nominativo, 1ª, masculino/feminino, plural: *Non sucedeu iso e nós entramos no goberno.*
- Rtn1fp→Pronome, tónico, nominativo, 1ª, feminino, plural: *E por enriba de todo elas, nós, tamén temos dereito a participar nas decisións.*
- Rtn1fs→Pronome, tónico, nominativo, 1ª, feminino, singular: *Pola miña parte, eu son mestra en ikebana polas escolas Sogetsu e Koriu, despois de sete anos de práctica no Xapón.*
- Rtn1mp→Pronome, tónico, nominativo, 1ª, masculino, plural: *Mar das conserveiras, dos armadores, que nos fixo cantar e contar aquilo de para mariñeiros nós.*
- Rtn1ms→Pronome, tónico, nominativo, 1ª, masculino, singular: *"Eu estou bastante afeito a escoita-los portavoces do PP no Parlamento loa-lo modo no que traballan..."*
- Rtn2as→Pronome, tónico, nominativo, 2ª, masculino/feminino, singular: *Por exemplo, se ti vas a un hotel facilmente te atoparás cun cadro moderno, non figurativo... é algo normal.*
- Rtn2ms→Pronome, tónico, nominativo, 2ª, masculino, singular: *Timbrio aqui, Febo alá, tirador acó, menciñeiro acolá, pai da Poesía, Inventor da Música, ti que decote alboreas e, aínda que o semella, endexamais te escureces!"*
- Rtn3ap→Pronome, tónico, nominativo, 3ª, masculino/feminino, plural: *"Fai falla dixo mante-lo equilibrio orzamentario para combater-la inflación, cousa que vostedes non queren;*
- Rtn3as→Pronome, tónico, nominativo, 3ª, masculino/feminino, singular: *Mais vostede non entrou.*
- Rtn3fp→Pronome, tónico, nominativo, 3ª, feminino, plural: *Todas elas saborosas dunha en unha e ao mesmo tempo susceptibles de ser mesturadas.*
- Rtn3fs→Pronome, tónico, nominativo, 3ª, feminino, singular: *Vostede era de procedencia galeguista, mais agora está integrada no PSdeG.*
- Rtn3mp→Pronome, tónico, nominativo, 3ª, masculino, plural: *O traballo non foi doado aínda que eles xa coñecían o sector da alimentación en Bélxica.*
- Rtn3ms→Pronome, tónico, nominativo, 3ª, masculino, singular: *El explicou que cómpre unha nova "xeografía comercial" que diversifique os intercambios.*
- Rao3aa→Pronome, átono, acusativo/dativo/impersoal, 3ª, masculino/feminino, singular/plural: *Hai unha parte do PT que se nega a acompañar o Lula e lle fai fronte.*

Rtp1ap→Pronome, tónico, preposicional, 1ª, masculino/feminino, plural: *A nós sorprendeunos o ben que soaba o Quixote en galego e o fácil que nos resultou collerlle o tono xeral.*

Rtp1as→Pronome, tónico, preposicional, 1ª, masculino/feminino, singular: *É un rango que ten o PSC que a min me gustaría que alcanzase o PSdeG.*

Rtp1fs→Pronome, tónico, preposicional, 1ª, feminino, singular: *Leo na revista CLIJ unhas palabras de Liliana Heker, escritora e directora do Fondo Nacional das Artes de Arxentina, nas que afirma que "a lectura, moitas veces, obrou en min sinxelamente como un refuxio e como un consolo".*

Rtp1mp→Pronome, tónico, preposicional, 1ª, masculino, plural: *"Para nosoutros é importante todo tipo de aforradores, incluso o dos 20.000 pesos".*

Rtp1ms→Pronome, tónico, preposicional, 1ª, masculino, singular: *Ese potencial de competitividade, esa audacia, é a que me fai sentir a min optimista como galego e como economista.*

Rtp2as→Pronome, tónico, preposicional, 2ª, masculino/feminino, singular: *Desde cando se rían de ti cando falabas galego a hoxe hai unha diferenza notábel.*

Rtp3aa→Pronome, tónico, preposicional, 3ª, masculino/feminino, singular/plural: *...os ataques terroristas do 11 de setembro apuntillaron un ano xa de por sí nefasto.*

Rtp3as→Pronome, tónico, preposicional, 3ª, masculino/feminino, singular: *_Polo que di, para vostede o éxito económico sobrepasa as razóns da produtividade.*

Rtp3fp→Pronome, tónico, preposicional, 3ª, feminino, plural: *é ilóxico intentar mercadear con elas.*

Rtp3fs→Pronome, tónico, preposicional, 3ª, feminino, singular: *Pilar lembrou que unha compañeira dela, Toñita, é desta parroquia e preguntoulle como poderíamos ver as pinturas.*

Rtp3mp→Pronome, tónico, preposicional, 3ª, masculino, plural: *O sistema do "sálvese quen poida" servirá para os que creen que poden vivir por sí mesmos lonxe de consensos custosos...*

Rtp3ms→Pronome, tónico, preposicional, 3ª, masculino, singular: *Que debemos esperar del?*

Rtp3ns→Pronome, tónico, preposicional, 3 neutro singular: *...favorecerá un aumento das vendas do sector editorial, e con elo, a difusión da cultura e o incentivo da creación literaria.*

SUBSTANTIVO (S) (14 etiquetas)

Scap→Substantivo, común, masculino/feminino, plural: *Cal considera a mellor escola para formar guitarristas?*

Scas→Substantivo, común, masculino/feminino, singular: *É tenente de alcalde de Pontevedra*

Scfp→Substantivo, común, feminino, plural: *Tamén se precisarían grabacións en audio da súa voz (lectura de poemas, conferencias, entrevistas...) e todo o que poda ser utilizado neste documental.*

Scfs→Substantivo, común, feminino, singular: *Alternativa contra a escasa rendibilidade da bolsa e do aforro*

Scma→Substantivo, común, masculino, singular/plural: *Esta empresa aproveita pezas de matadoiro, entre elas páncreas para insulina;*

Scmp→Substantivo, común, masculino, plural: *A causa: a alta tendencia inflacionista, as escasas marxes de beneficios, as cotas de mantemento e os impostos.*

Scms→Substantivo, común, masculino, singular: *O ingreso de España no proceso de implantación da moeda única europea é un acontecemento da maior importancia.*

Sp00→Substantivo, propio, xénero non aplica, número non aplica: *Así funcionan as cousas en Vigo.*

Spf0→Substantivo, propio, feminino, número non aplica: *Ana Isabel García Arias / Pilar lembrou que unha compañeira dela, Toñita, é desta parroquia e preguntoulle como poderíamos ver as pinturas.*

Spfp→Substantivo, propio, feminino, plural: *O edificio, de estilo sobrio e funcional, pecha as catro frontes que escoltan a fonte das Praterías.*

Spfs→Substantivo, propio, feminino, singular: *Creo nunha España plural, pero non quero fronteiras.*

Spm0→Substantivo, propio, masculino, número non aplica: *César Gómez explica que o invento xa está patentado...*

Spmp→Substantivo, propio, masculino, plural: *Os Presupostos do Estado para o 2003 parecen feitos por Harry Potter.*

Spms→Substantivo, propio, masculino, singular: *A república do Xacobeo foi como a versión invertida do conto do Rei Midas, e quen tocaba a Sociedade de Xestión conseguía un pouco de ouro:*

RELATIVO (T) (11 etiquetas)

Tdfp→Relativo, determinante, feminino, plural: *...hai compañeiros sen contrato ou que traballan máis horas das que lle corresponden ou fotografías cuxas imaxes son usadas pola empresa sen recoñecerlle os dereitos de autor.*

Tdfs→Relativo, determinante, feminino, singular: *Unha Constitución cuxa reforma vai aprobar xa o Consello de Ministros esta semana, para fixar, deste xeito os límites dos estatutos.*

Tdmp→Relativo, determinante, masculino, plural: *...quedou excluída da rede enerxética prioritaria da UE, cuxos proxectos recibirán 28.000 millóns de euros.*

Tdms→Relativo, determinante, masculino, singular: *...concluíu onte un xuíz de instrución, cuxo ditame se coñeceu tras unha campaña da súa viúva para esclarecer a traxedia.*

Tnaa→Relativo, non determinante, masculino/feminino, singular/plural: *Ou, máis ben, que as autoridades centrais e autonómicas aínda non foron quen de transmitir as peculiaridades que están a impedir que Galicia se beneficie...*

Tnap→Relativo, non determinante, masculino/feminino, plural: *...ou dar traballo ás mozas e mozos que xa non emigran, ou mil outras cousas máis?*

Tnas→Relativo, non determinante, masculino/feminino, singular: *Seoane, que ofreceu unha roda de prensa para valorar a decisión de Política Territorial de non lle outorgar a citada aprobación debido a deficiencias...*

Tnfp→Relativo, non determinante, feminino, plural: *"Estas partidas que se repiten refírense sobre todo a obras menores..."*

Tnfs→Relativo, non determinante, feminino, singular: *Kafloat é o nome da empresa coruñesa que comercializará internacionalmente un sistema para evitar que os barcos naufraguen cando se produza unha entrada de auga.*

Tnmp→Relativo, non determinante, masculino, plural: *"Haberá membros do equipo directivo que permanezan aquí e membros do equipo directivo que estarán alí".*

Tnms→Relativo, non determinante, masculino, singular: *O novo consello de administración de Fadesa, que o xoves tomou posesión na cidade da Coruña, xa está ó choio.*

VERBO (V) (55 etiquetas)

V0f000→Verbo, tempo non aplica, infinitivo, persoa non aplica, xénero non aplica, número non aplica: *Din del que é o home que fixo soñar a De Gaulle e Miterrand con que Francia podería converterse nunha potencia militar e espacial.*

V0f10p→Verbo, tempo non aplica, infinitivo, 1ª, xénero non aplica, plural: *Ás veces podemos escoller para andar pés latinos ou gregos: sermos bípedes ou dípodos, e irmos ao pedicuro ou ao podólogo.*

V0f30p→Verbo, tempo non aplica, infinitivo, 3ª, xénero non aplica, plural: *Para o financeiro, a pesar de se cumpriren os prazos, é preocupante o futuro de sectores como a pesca ou o turismo...*

V0m20p→Verbo, tempo non aplica, imperativo, 2ª, xénero non aplica, plural: *Pensade en que actividades facedes habitualmente e analizar se as realizades de forma correcta, e considerade tamén o importante que é durmir nun bon colchón...*

V0m20s→Verbo, tempo non aplica, imperativo, 2ª, xénero non aplica, singular: *Mira que mono queda, cinco homes e cinco mulleres;*

V0p0fp→Verbo, tempo non aplica, participio, persoa non aplica, feminino, plural: *... xa que hai numerosas obras que terán que ser co-financiadas pola Xunta".*

V0p0fs→Verbo, tempo non aplica, participio, persoa non aplica, feminino, singular: *Unha vez demostrada a validez do invento, a sociedade comezará coa busca dos materiais...*

V0p0mp→Verbo, tempo non aplica, participio, persoa non aplica, masculino, plural: *Estes impostos compartidos deberían, cando menos, ser xestionados por unha Axencia Tributaria Común...*

V0p0ms→Verbo, tempo non aplica, participio, persoa non aplica, masculino, singular: *Dado o tamaño das compañías galegas, esta acción podería suxerir o recurso ó corporativismo...*

V0x000→Verbo, tempo non aplica, xerundio, persoa non aplica, xénero non aplica, número non aplica:
Os débitos a clientes, excluindo o efecto da cesión temporal de activos, presentan un crecemento...

V0x10p→Verbo, tempo non aplica, xerundio, 1ª, xénero non aplica, plural: *Xa saben, Sartre defendía faládomos de literatura _ que o escritor debe estar comprometido co seu tempo:*

Vci10p→Verbo, pospretérito, indicativo, 1ª, xénero non aplica, plural: *"Con esta medida conseguiríamos unha tributación moito máis mesurada, e aumentariámo-la recadación...*

Vci10s→Verbo, pospretérito, indicativo, 1ª, xénero non aplica, singular: *Centrándonos no mundo que nos rodea máis de cerca, eu diría que o fundamento está na xestión.*

Vci30p→Verbo, pospretérito, indicativo, 3ª, xénero non aplica, plural: *Tras el poderían vir novas incorporacións, xa que o nome de Galiza vai facéndose un espacio entre os consumidores...*

Vci30s→Verbo, pospretérito, indicativo, 3ª, xénero non aplica, singular: *O primeiro sería crear institucións supranacionais con capacidade política.*

Vcia0s→Verbo, pospretérito, indicativo, 1ª/3ª, xénero non aplica, singular: *Diría máis:*

Vei10p→Verbo, pretérito, indicativo, 1ª, xénero non aplica, plural: *Lembro por exemplo que fixemos un esforzo por buscar a versión galega de todos os refráns que usou.*

Vei10s→Verbo, pretérito, indicativo, 1ª, xénero non aplica, singular: *Eu dei varios cursiños de ikebana nas asociacións culturais de Vidán e de Vite, e decateime de que a xente desfruta aprendéndoa.*

Vei20p→Verbo, pretérito, indicativo, 2ª, xénero non aplica, plural: *Nalgún festival o ska xa é máis que un toque decorativo, hai pouco tocáchedes no Sagarrandotik...*

Vei20s→Verbo, pretérito, indicativo, 2ª, xénero non aplica, singular: *"Explicalle aos da rúa o que viches aquí" a frase coa que se despide dela unha das reclusas...*

Vei30p→Verbo, pretérito, indicativo, 3ª, xénero non aplica, plural: *É o caso da autovía transcantábrica, da que se presupostaron dous millóns de euros.*

Vei30s→Verbo, pretérito, indicativo, 3ª, xénero non aplica, singular: *A idea xurdiu hai catro anos da man de dous empresarios da Coruña, Vicente López e César Gómez.*

Vfi10p→Verbo, futuro, indicativo, 1ª, xénero non aplica, plural: *Conseguiremos unha sociedade sen medos e audaz.*

Vfi10s→Verbo, futuro, indicativo, 1ª, xénero non aplica, singular: *De seguido, tentarei facer unha síntese deste Memorandum.*

Vfi20p→Verbo, futuro, indicativo, 2ª, xénero non aplica, plural: *"Agora que acabo de ser elixido presidente da República e en vésperas de asumir o cargo, todos comprenderedes que non poida seguir á cabeza" do partido, dixo Sarkozy,...*

- Vfi20s→Verbo, futuro, indicativo, 2ª, xénero non aplica, singular: *Por exemplo, se ti vas a un hotel facilmente te atoparás cun cadro moderno, non figurativo... é algo normal.*
- Vfi30p→Verbo, futuro, indicativo, 3ª, xénero non aplica, plural: *En total serán 500,72 millóns de euros os fondos públicos que gastarán os distintos Ministerios na nosa comunidade, dos que case 100 millóns non deberían ser incluídos.*
- Vfi30s→Verbo, futuro, indicativo, 3ª, xénero non aplica, singular: *O mecanismo comercializarase no 2003*
- Vfs30s→Verbo, futuro, subxuntivo, 3ª, xénero non aplica, singular: *Toca agora, partindo deste traballo, elaborar un texto articulado quen de se apresentar no Parlamento galego e de se confrontar co que propoñan _se tal for a súa vontade_ as demais forzas políticas actúantes na Galiza.*
- Ves10p→Verbo, pretérito, subxuntivo, 1ª, xénero non aplica, plural: *"Se non vendeseamos ese cuarzo teríamos un gran vertedoiro", di Goyoaga.*
- Ves10s→Verbo, pretérito, subxuntivo, 1ª, xénero non aplica, singular: *Se eu fose alcalde da Coruña o primeiro que faría sería solucionar o problema do tráfico.*
- Ves30p→Verbo, pretérito, subxuntivo, 3ª, xénero non aplica, plural: *Esas expresións que moi axiña nos parece imposíbel que non tivesen existido antes.*
- Ves30s→Verbo, pretérito, subxuntivo, 3ª, xénero non aplica, singular: *Non era necesario e foi suficiente cunha insinuación de Riera para que a alcaldesa fixese mutis e desaparecese da escena.*
- Vii10p→Verbo, copretérito, indicativo, 1ª, xénero non aplica, plural: *Non entendo por que os homes eramos discriminados.*
- Vii10s→Verbo, copretérito, indicativo, 1ª, xénero non aplica, singular: *"O que tiña no corpo levaba 12 anos aí metido sen saber que o tiña", explicou Mari Carmen Fernández...*
- Vii20s→Verbo, copretérito, indicativo, 2ª, xénero non aplica, singular: *Desde cando se rían de ti cando falabas galego a hoxe hai unha diferenza notábel.*
- Vii30p→Verbo, copretérito, indicativo, 3ª, xénero non aplica, plural: *Os estaleiros de Ferrol, que ocupaban arredor de 14.000 traballadores cando corrían os anos 70, fican agora atendidos por 2.000, na súa maior parte administrativos e técnicos.*
- Vii30s→Verbo, copretérito, indicativo, 3ª, xénero non aplica, singular: *Era o cantar da miña nai que eu máis gozaba de ouvir: forza, graza, alegría e raiba para bailar.*
- Via0s→Verbo, copretérito, indicativo, 1ª/3ª, xénero non aplica, singular: *Pensaba estes días en que o oasis político catalán se ve ameazado por unha dura treboada de area, desas capaces de deixar tras o amencer unha árida paisaxe.*

- Vli10p→Verbo, antepretérito, indicativo, 1ª, xénero non aplica, plural: *...versións ambas que se din respectuosas destes cantos que nunca debéramos dar por esquecidos, pois neles está recollida a esencia do noso hábitat.*
- Vli10s→Verbo, antepretérito, indicativo, 1ª, xénero non aplica, singular: *Dito isto quixera formular tres cuestións:*
- Vli30p→Verbo, antepretérito, indicativo, 3ª, xénero non aplica, plural: *E aquela viaxe rematara como lle estaban a indicar os deuses que escribiran a historia inconclusa do Home";*
- Vli30s→Verbo, antepretérito, indicativo, 3ª, xénero non aplica, singular: *Os ollos da ponte, de Ramiro Fonte, volve ao escenario que o escritor eumés reflectira en Os meus ollos.*
- Vpi10p→Verbo, presente, indicativo, 1ª, xénero non aplica, plural: *Ollando cara diante, só vemos que o problema se vai agravar.*
- Vpi10s→Verbo, presente, indicativo, 1ª, xénero non aplica, singular: *Sigo a selo.*
- Vpi20p→Verbo, presente, indicativo, 2ª, xénero non aplica, plural: *Se tedes en consideración estes factores será máis difícil autodanarse.*
- Vpi20s→Verbo, presente, indicativo, 2ª, xénero non aplica, singular: *E sentes que te alongas voando cara ao horizonte".*
- Vpi30p→Verbo, presente, indicativo, 3ª, xénero non aplica, plural: *Se as políticas do Fondo Monetario para Arxentina se levan a cabo tal e como se están a apresentar agora, Arxentina desintegrarase.*
- Vpi30s→Verbo, presente, indicativo, 3ª, xénero non aplica, singular: *Esta cifra supón unha media de 283 euros por habitante, fronte os 392 do conxunto do Estado.*
- Vps10p→Verbo, presente, subxuntivo, 1ª, xénero non aplica, plural: *Coido que vai chegando o intre de que os galegos lle erixamos unha estatua ó Sr. Jordi Pujol.*
- Vps10s→Verbo, presente, subxuntivo, 1ª, xénero non aplica, singular: *Pode que perda a batalla nesa maneira de interpretar competitivamente o mundo, pero vouno intentar.*
- Vps20s→Verbo, presente, subxuntivo, 2ª, xénero non aplica, singular: *É raro que escoites música atonal en cine ou na publicidade, e por iso non hai avance.*
- Vps30p→Verbo, presente, subxuntivo, 3ª, xénero non aplica, plural: *"A cultura da pesca que teñen os galegos comenta Paz-Andrade fai que coiden o mar como o fai pouca xente.*
- Vps30s→Verbo, presente, subxuntivo, 3ª, xénero non aplica, singular: *"Non creo que Europa este sendo boa para nós, pero non podó afirmar con total seguridade que é mala".*
- Vpsa0s→Verbo, presente, subxuntivo, 1ª/3ª, xénero non aplica, singular: *Pero diso non se pode deducir que, en canto ó calendario, avogue por un aprazamento, pois non é así en absoluto", subliñou.*

ADVERBIO (W) (4 etiquetas)

Wg→Adverbio, interrogativo/exclamativo: *O chanceler alemán, Helmut Kohl, declarábase onte, sábado, canso de discusións sobre cómo, cándo e con quén comezará a unión económica e monetaria europea (UEM), e recomendoulles a tódolos países da UE que fagan os seus deberes.*

Wm→Adverbio, modificador: *A Coordinadora Galega de ONGD valora moi positivamente este proxecto e o proceso de xestión que o ten precedido.*

Wn→Adverbio, nuclear: *"Obviamente non é o mesmo manter aboiando e mover unha tonelada que dez.*

Wr→Adverbio, relativo: *Neste sentido, o que desexa a patronal é que as tarifas estean ligadas ós custos de produción de forma similar a como fluctúan os prezos dos petrolíferos por parte dos operadores do mercado.*

INTERXECCIÓN (y) (1 etiqueta)

Y→Interxección: *Están moi equivocados, claro. / Bueno, financiamos proxectos xornalísticos, concedemos créditos a xornais, iso sí... / As comunidades máis mediterráneas e Madrid son, ¡ai!, contribuíntes netos. / ¡Meu Deus!*

PERIFÉRICA (Z) (12 etiquetas)

Za00→Periférica, abreviatura, xénero non aplica, número non aplica: *A desaparición da peseta, a pela, a rubia, etc., non parece que nos poña moi nostáxicos.*

Zafp→Periférica, abreviatura, feminino, plural: *Por exemplo, en Galicia supón ao redor de 75.000 millóns de ptas.*

Zafs→Periférica, abreviatura, feminino, singular: *O socialista Joaquín Almunia, (da), xunto a Txiqui Benegas*

Zamp→Periférica, abreviatura, masculino, plural: *"Termino Sres. Accionistas.*

Zams→Periférica, abreviatura, masculino, singular: *O 4M debería se-lo gran día, sexa quen sexa o que deba gobernar:*

Zg00→Periférica, sigla, xénero non aplica, número non aplica: *Como máxima dirixente local dos socialistas manifesta unha posición inequívoca a prol do traslado de ENCE.*

Zgfp→Periférica, sigla, feminino, plural: *Á marxe, as APA e organizacións de consumidores parecen satisfeitas coa resolución...*

Zgfs→Periférica, sigla, feminino, singular: *González Viñas asegura que a SEPI diría que si ó seu proxecto se non existise a prohibición de Bruxelas*

Zgmp→Periférica, sigla, masculino, plural: *Na denominada operación "Réplica-rede" efectuáronse un total de 20 rexistros domiciliarios e interviñéronse 32 discos duros, dous ordenadores portátiles e numerosos Dvd con máis de 10.000 imaxes...*

Zgms→Periférica, sigla, masculino, singular: *Corina Porro volve ós ataques a Caride polo PXOM de Vigo.*

Zo00→Periférica, outros, xénero non aplica, número non aplica: *www.xeracionweb.com; 7147BRB; C); 2.; 901 222250...*

Zs00→Periférica, símbolo, xénero non aplica, número non aplica: *pts./litro, kwh, home/vehículo, CO2, kg, MW, Hl/ha...*

Apéndice C

Listaxe de etiquetas recollidas no kernel pero non documentadas na versión 2.1 do CORGAetq

A listaxe que presentamos deseguido recolle as etiquetas que aparecen no *kernel* do corpus de adestramento do etiquetador de XIADA e que non se documentaron, polo momento, no corpus etiquetado automaticamente e desambiguado á man (CORGAetq, versión 2.1, que consta de 386.498 formas ortográficas ou en elementos gramaticais, 426.264). A relación que ofrecemos está ordenada alfabeticamente pola etiqueta, a continuación da cal se desenvolve o seu valor e se engade un exemplo extraído do núcleo ou *kernel* de XIADA etiquetado, aquel no que se recollen todas as etiquetas posibles, polo menos, unha vez.

Consta esta relación de 84 etiquetas, que xunto coas 299 do apéndice anterior (383 en total), dan conta, en teoría, da caracterización morfolóxica de calquera palabra pertencente á lingua galega e, na práctica, permiten a adecuación desa caracterización a un contexto dado.

ADXECTIVO (A) (5 etiquetas)

A0fa→Adxectivo, grao non aplica, feminino, singular/plural: ¡*Sífilis mexericas!*

A0ma→Adxectivo, grao non aplica, masculino, singular/plural: ¡*Parabrisas furafollas!*

Acap→Adxectivo, comparativo, masculino/feminino, plural: *Convoca a proba libre para a obtención do título de graduado en educación secundaria para maiores de 18 anos na convocatoria de xuño e de setembro e se dictan instrucións para realizala.*

Asap→Adxectivo, superlativo, masculino/feminino, plural: *Son hipersensibles á luz.*

Asas→Adxectivo, superlativo, masculino/feminino, singular: *Ofenderíase moi profundamente, pois neste punto é hipersensibel*

INTERROGATIVO-EXCLAMATIVO (C) (3 etiquetas)

Gdap→Interrogativo-exclamativo, determinante, masculino/feminino, plural: ¡*Que fáciles!*

Gdas→Interrogativo-exclamativo, determinante, masculino/feminino, singular: *Que terrible.*

Gnap→Interrogativo-exclamativo, non determinante, masculino/feminino, plural: ¡*E quen veñen á Mesa de Programación?*

INDEFINIDO (I) (1 etiqueta)

Inap→Indefinido, non determinante, masculino/feminino, plural: *Ninguén sospeitabamos que as medidas a tomar foran xa as tan mortuorias dos nichos de Bonaval.*

POSESIVO (M) (18 etiquetas)

Md2pfp→Posesivo, determinante, 2ª, plural, feminino, plural: *Cos teus cans, as túas gatas, os vosos cabalos, as vosas vacas, o voso elefante e as miñas galiñas xa teriades unha boa granxa se quixesedes montala.*

Md2sfp→Posesivo, determinante, 2ª, singular, feminino, plural: *Cos teus cans, as túas gatas, os vosos cabalos, as vosas vacas, o voso elefante e as miñas galiñas xa teriades unha boa granxa se quixesedes montala.*

Mn1pfp→Posesivo, non determinante, 1ª, plural, feminino, plural: *Reivindicaron que as rúas "tamén son nosas, non só das multinacionais", ás que acusaron de aproveita-las condicións nas que se encontran os países do terceiro mundo para incrementa-los seus beneficios.*

Mn1pmp→Posesivo, non determinante, 1ª, plural, masculino, plural: *Dos nosos, o que en peor situación se encontraba era o Silverio Gándara, que seguía con aquelas chagas das pernas, que daba mágoa verllas.*

Mn2pfp→Posesivo, non determinante, 2ª, plural, feminino, plural: *A min tanto me ten, aínda que xa che dixen que non teño nada que ver coas pantasma vosas.*

Mn2pfs→Posesivo, non determinante, 2ª, plural, feminino, singular: *Eu non sei, pero a nosa raza é diferente da vosa.*

Mn2pmp→Posesivo, non determinante, 2ª, plural, masculino, plural: *_Mira a ver se anda por aí algún dos teus, e dormes nun camión dos vosos, ho.*

Mn2pms→Posesivo, non determinante, 2ª, plural, masculino, singular: *O home que está aquí _e sinalou agora ó Candal _ non quere matarvos, quere ser amigo vosso.*

Mn2sfp→Posesivo, non determinante, 2ª, singular, feminino, plural: *Son figuracións túas.*

Mn2sfs→Posesivo, non determinante, 2ª, singular, feminino, singular: *Resulta evidente que unha noite en soedade é moito máis noite e tamén é moita máis soedade, un séntese impotente diante da escuridade aberta e sincera, como diante túa.*

Mn2smp→Posesivo, non determinante, 2ª, singular, masculino, plural: *_Mira a ver se anda por aí algún dos teus, e dormes nun camión dos vosos, ho.*

Mn2sms→Posesivo, non determinante, 2ª, singular, masculino, singular: *_¿Aínda te emocionas cando sae un libro teu?*

Mn3afp→Posesivo, non determinante, 3ª, singular/plural, feminino, plural: *Entre a oficialidade murmurábase que tal e tal casa eran súas.*

Mn3afs→Posesivo, non determinante, 3ª, singular/plural, feminino, singular: *Aquela casa xa non era súa, era unha cela negra de paredes difusas, un espacio equívoco e sen termo, a superficie inabarcable e informe da súa limitación.*

Mn3amp→Posesivo, non determinante, 3ª, singular/plural, masculino, plural: *Eses contos son seus?*

Mn3ams→Posesivo, non determinante, 3ª, singular/plural, masculino, singular: *E escoitar non é o seu.*

Mn3pfp→Posesivo, non determinante, 3ª, plural, feminino, plural: *Fun ao parque e pouseinas ao pé da fonte, onde eles lavan as súas.*

Mn3pmp→Posesivo, non determinante, 3ª, plural, masculino, plural: *Dixeran que eran seus.*

SIGNO DE PUNTUACIÓN (Q) (2 etiquetas)

Q{→Signo de puntuación {→*Regaleilles {no meu aniversario} cadanseu anel e cadansúas bonecas aos vosos netos e eles agasalláronme con cadanseus panos e cadansúa revista.*

Q}→Signo de puntuación }

PRONOME (R) (27 etiquetas)

Raa2fp→Pronome, átono, acusativo, 2ª, feminino, plural: *Ademais, non vos vou deixar a vós soias co rapaz.*

Raa2fs→Pronome, átono, acusativo, 2ª, feminino, singular: *Non sei nin como ti te atreves a andar así, soliña pola vila adiante, para que te rapten e despois lle vaian poñer condicións ó teu pai.*

Raa2mp→Pronome, átono, acusativo, 2ª, masculino, plural: *"Tamén a vós vos terei que deixar, meus amigos pensou.*

Rad1fp→Pronome, átono, dativo, 1ª, feminino, plural: *JPois xa ves a nós as dúas de que nos serviu!*

Rad2fp→Pronome, átono, dativo, 2ª, feminino, plural: *A vós, as da comunidade, faivos moita falta un home con experiencia, que vos poña a andar .*

Rad2fs→Pronome, átono, dativo, 2ª, feminino, singular: *Para ti, Berta, as mulleres... témosche todas que ser dun mesmo estilo.*

Rad2mp→Pronome, átono, dativo, 2ª, masculino, plural: *Xa volo contarei ós dous xuntos desculpouse.*

Rad2ms→Pronome, átono, dativo, 2ª, masculino, singular: *En fin, ¿quén che di a ti, Cibrán, que a Marisa, acabado o bachillerato, non estudiaba medicina ela mesma, ou outra cousa nova desas que hai agora?*

Raf1fp→Pronome, átono, formante léxico, 1ª, feminino, plural: *Atreverémonos a facelo nó-las dúas?*

Raf1fs→Pronome, átono, formante léxico, 1ª, feminino, singular: *Acórdome de pequena, todos os doces que nos compraba, pero volveuse insoportable desque a chegada dos socialistas ó poder detivo a súa carreira no Exército.*

Raf2ap→Pronome, átono, formante léxico, 2ª, masculino/feminino, plural: *¿Xa non vos acordades?*

Raf2fp→Pronome, átono, formante léxico, 2ª, feminino, plural: *Neniñas, non debes espreguizarvos diante de todos!*

Raf2fs→Pronome, átono, formante léxico, 2ª, feminino, singular: *Non sei nin como ti te atreves a andar así, soliña pola vila adiante, para que te rapten e despois lle vaian poñer condicións ó teu pai.*

Raf2mp→Pronome, átono, formante léxico, 2ª, masculino, plural: *"Fóra, fóra, veña, a fumar des na casiña diante dos vosos pais, desgraciados, a ver se vos atrevedes", e cando saían os rapaces seguía o Evaristo:*

Raf2ms→Pronome, átono, formante léxico, 2ª, masculino, singular: *Acórdate de min, Carlos, acórdate de min e de todo.*

Rtn2ap→Pronome, tónico, nominativo, 2ª, masculino/feminino, plural: *Nós xa chegamos e vós xa comerades cando el comezou a falar.*

Rtn2fp→Pronome, tónico, nominativo, 2ª, feminino, plural: *Vosoutras así o quixestes.*

Rtn2fs→Pronome, tónico, nominativo, 2ª, feminino, singular: *Ti mesma repetíchelo insistentemente.*

Rtn2mp→Pronome, tónico, nominativo, 2ª, masculino, plural: *E vosoutros dixo, dirixíndose aos pícaros, ben se ve que hoxe non tendes escola.*

Rtn3ns→Pronome, tónico, nominativo, 3ª, neutro, singular: *Tamén é un medio bo para combatilo, fomentar as nosas pequenas ambicións, loitar para realizalas cabalmente e amar o esforzo que elo require.*

Rtp1fp→Pronome, tónico, preposicional, 1ª, feminino, plural: *Non falou con vosoutros nin con nosoutras.*

Rtp2ap→Pronome, tónico, preposicional, 2ª, masculino/feminino, plural: *No "Biqueque" quería vervos eu a vós.*

Rtp2fp→Pronome, tónico, preposicional, 2ª, feminino, plural: *¿Conto con vosoutras?*

Rtp2fs→Pronome, tónico, preposicional, 2ª, feminino, singular: *¿E quen che di a ti, Elvira, que a túa enfermidade vai continuar así, que non podes mellorar?*

Rtp2mp→Pronome, tónico, preposicional, 2ª, masculino, plural: *Non falou con vosoutros nin con nosoutras.*

Rtp2ms→Pronome, tónico, preposicional, 2ª, masculino, singular: *Nin para ti nin para ninguén, fillo.*

Rtp3ap→Pronome, tónico, preposicional, 3ª, masculino/feminino, plural: *Realmente, foi un milagre, un verdadeiro milagre, que non deran con vostedes.*

SUBSTANTIVO (S) (2 etiquetas)

Scaa→Substantivo, común, masculino/feminino, singular/plural: *¿Que hai, matasáns?*

Scfa→Substantivo, común, feminino, singular/plural: *¡Sífilis mexericas!*

VERBO (V) (20 etiquetas)

V0f20p→Verbo, tempo non aplica, infinitivo, 2ª, xénero non aplica, plural: *Pagardes entre todos ou pagares ti non semella tan difícil decisión.*

V0f20s→Verbo, tempo non aplica, infinitivo, 2ª, xénero non aplica, singular: *Pagardes entre todos ou pagares ti non semella tan difícil decisión.*

V0m10p→Verbo, tempo non aplica, imperativo, 1ª, xénero non aplica, plural: *¡Vamos apuntarnos! _contestou Xurxo.*

V0x20p→Verbo, tempo non aplica, xerundio, 2ª, xénero non aplica, plural: *Creo que todo sairá mellor facéndodes como di el.*

Vci20p→Verbo, pospretérito, indicativo, 2ª, xénero non aplica, plural: *Cos teus cans, as túas gatas, os vosos cabalos, as vosas vacas, o voso elefante e as miñas galiñas xa teriades unha boa granxa se quixesedes montala.*

Vci20s→Verbo, pospretérito, indicativo, 2ª, xénero non aplica, singular: *Contases ti ou contaramos nós, enganaríaste de tódolos xeitos.*

Ves20p→Verbo, pretérito, subxuntivo, 2ª, xénero non aplica, plural: *Cos teus cans, as túas gatas, os vosos cabalos, as vosas vacas, o voso elefante e as miñas galiñas xa teriades unha boa granxa se quixesedes montala.*

Ves20s→Verbo, pretérito, subxuntivo, 2ª, xénero non aplica, singular: *Contases ti ou contaramos nós, enganaríaste de tódolos xeitos.*

Vesa0s→Verbo, pretérito, subxuntivo, 1ª/3ª, xénero non aplica, singular: *A cuestión non é que bailase ben senón que cantaba moi mal.*

Vfs10p→Verbo, futuro, subxuntivo, 1ª, xénero non aplica, plural: *Como sementarmos, así colleremos cando seguemos.*

Vfs10s→Verbo, futuro, subxuntivo, 1ª, xénero non aplica, singular: *Xuro que se non morrer aquí hei volver.*

Vfs20p→Verbo, futuro, subxuntivo, 2ª, xénero non aplica, plural: *Onde virdes moita comida non vos metades.*

Vfs20s→Verbo, futuro, subxuntivo, 2ª, xénero non aplica, singular: *Onde vires moito fume non te vaias quentar.*

Vfs30p→Verbo, futuro, subxuntivo, 3ª, xénero non aplica, plural: *Mais se en xuicio conservou o comercio en pratas e cobre, non tardaría en perdelo polo seu mal tino prao negocio, que todo era chegaren donceis lisboetas e malvenderlles todo o que eles quixeren.*

Vfsa0s→Verbo, futuro, subxuntivo, 1ª/3ª, xénero non aplica, singular: *Iso foi por fixer todo o traballo.*

Vii20p→Verbo, copretérito, indicativo, 2ª, xénero non aplica, plural: *O crego non falou convosco porque non estabades na casa, pero si falou contigo posto que estabas e non che deu a razón.*

Vli20p→Verbo, antepretérito, indicativo, 2ª, xénero non aplica, plural: *Nós xa chegamos e vós xa comerades cando el comezou a falar.*

Vli20s→Verbo, antepretérito, indicativo, 2ª, xénero non aplica, singular: *Non fai máis ca repetir que aquela noite estiveras encantadora.*

Vlia0s→Verbo, antepretérito, indicativo, 1ª/3ª, xénero non aplica, singular: *Dixo que o fixera moi ben.*

Vps20p→Verbo, presente, subxuntivo, 2ª, xénero non aplica, plural: *Onde virdes moita comida non vos metades.*

CATEGORÍA PERIFÉRICA (Z) (6 etiquetas)

Zaap→Periférica, abreviatura, masculino/feminino, plural: *O Dr. Silva, a dra. Ferro, os Sres. Martínez e Gómez e as Sras. María García e Olaia Baleato dirixense a Vdes. para presentarlles un ensaio sobre atlas [o acto terá lugar no Salón Principal].*

Zaas→Periférica, abreviatura, masculino/feminino, singular: *Por moito que a Vde. lle amole.*

Zf00→Periférica, fórmula, xénero non aplica, número non aplica: *Supoñamos na superficie da Terra unha presión P = 101 300 Pa.*

Zgaa→Periférica, abreviatura, masculino/feminino, singular/plural: *En concreto, Sanidade oferta 170 prazas do grupo de auxiliar administrativo, 165 de celadores, 150 de auxiliar de enfermería, 115 de ATS, 51 de asistentes sociais, 25 de técnicos especialistas de laboratorio, 24 de grupo técnico da función administrativa, 21 de técnicos especialistas de radioloxía, 21 de pinches e 20 de psicólogos.*

Zgfa→Periférica, sigla, feminino, singular/plural: *Logo, en canto á frecuencia de aparición nos informativos encóntranse na tese os 'elementos anónimos', que se relacionan cos sucesos, e, finalmente, os 'novos motores sociais', en alusión a entidades como ONG, etc.*

Zgma→Periférica, sigla, masculino, singular/plural: *Esta comisión comezará a traballar en xaneiro, se ben non será ata principios de marzo cando se elixa no Claustro a representación de profesores, alumnos e PAS.*

Apéndice D

Listaxe de etiquetas usadas só no lexicón de Xiada

A listaxe que aquí presentamos recolle soamente etiquetas que se utilizaron ou utilizan aínda para a incorporación dalgúns elementos ao lexicón de XIADA e que non aparecen nin poden aparecer na etiquetación de corpus, pois non responden á análise lingüística real do galego senón que son froito dunha solución práctica adoptada para introducir o léxico na base de datos.

Excepto as tres verbais, que se explican no capítulo dedicado á desambiguación, as restantes 67 débense á aglutinación de valores nun único carácter co fin de facilitar a introdución no lexicón de certos elementos de categorías pechadas ou dos adverbios. Por exemplo, para evitar ter que introducir no lexicón dúas entradas por cada un dos adverbios que poden funcionar como *modificadores* (*Wm =Adverbio modificador*) ou *nucleares* (*Wn =Adverbio nuclear*), creamos un valor “a” (*Wa =Adverbio nuclear/modificador*) que nos permite incorporar cunha única entrada os dous valores, xa que, logo, á hora de extraer o léxico ou aplicar o etiquetario á etiquetación de corpus, o responsable informático de XIADA converte ese valor inventado (*Wa*) nos dous únicos reais: *Wn (Adverbio nuclear)* e *Wm (Adverbio modificador)*.

Ademais de no Adverbio, a indefinición ou aglutinación dos valores do atributo *valor* dáse nalgúns elementos das categorías Demostrativo, Interrogativo-exclamativo, Indefinido, Posesivo, Numeral e Relativo, coa consecuente amalgama dos valores *determinante/non determinante*; no Pronome os que se agrupan no lexicón son os valores casuais *nominativo/preposicional* e, finalmente, nas formas do posesivo de respecto xúntase o valor xenérico *masculino/feminino*.

A relación que ofrecemos (70 etiquetas) está ordenada alfabeticamente pola etiqueta, a continuación da cal se desenvolve o seu valor e se engade un ou varios exemplos.

DEMOSTRATIVO (E) (4 etiquetas)

Eafp→Demostrativo, determinante/non determinante, feminino, plural: *estas, esas, aquelas*

Eafs→Demostrativo, determinante/non determinante, feminino, singular: *esta, esa, aquela*

Eamp→Demostrativo, determinante/non determinante, masculino, plural: *estes, eses, aqueles*

Eams→Demostrativo, determinante/non determinante, masculino, singular: *este, ese, aquel*

INTERROGATIVO-EXCLAMATIVO (G) (7 etiquetas)

Gaaa→Interrogativo-exclamativo, determinante/non determinante, masculino/feminino, singular/plural:
que

Gaap→Interrogativo-exclamativo, determinante/non determinante, masculino/feminino, plural: *cales*

Gaas→Interrogativo-exclamativo, determinante/non determinante, masculino/feminino, singular: *cal*

Gafp→Interrogativo-exclamativo, determinante/non determinante, feminino, plural: *cantas*

Gafs→Interrogativo-exclamativo, determinante/non determinante, feminino, singular: *canta*

Gamp→Interrogativo-exclamativo, determinante/non determinante, masculino, plural: *cantos*

Gams→Interrogativo-exclamativo, determinante/non determinante, masculino, singular: *canto*

INDEFINIDO (I) (3 etiquetas)

Iaaa→Indefinido, determinante/non determinante, masculino/feminino, singular/plural: *calquera, demais*

Iaap→Indefinido, determinante/non determinante, masculino/feminino, plural: *tales*

Iaas→Indefinido, determinante/non determinante, masculino/feminino, singular: *tal*

POSESIVO (E) (26 etiquetas)

Ma1pfp→Posesivo, determinante/non determinante, 1ª, plural, feminino, plural: *nosas*

Ma1pfs→Posesivo, determinante/non determinante, 1ª, plural, feminino, singular: *nosa*

Ma1pmp→Posesivo, determinante/non determinante, 1ª, plural, masculino, plural: *nosos*

Ma1pms→Posesivo, determinante/non determinante, 1ª, plural, masculino, singular: *noso*

Ma1sfp→Posesivo, determinante/non determinante, 1ª, singular, feminino, plural: *miñas*

Ma1sfs→Posesivo, determinante/non determinante, 1ª, singular, feminino, singular: *miña*

Ma1smp→Posesivo, determinante/non determinante, 1ª, singular, masculino, plural: *meus*

Ma1sms→Posesivo, determinante/non determinante, 1ª, singular, masculino, singular: *meu*

Ma2pfp→Posesivo, determinante/non determinante, 2ª, plural, feminino, plural: *vosas*

Ma2pfs→Posesivo, determinante/non determinante, 2ª, plural, feminino, singular: *vosa*

Ma2pmp→Posesivo, determinante/non determinante, 2ª, plural, masculino, plural: *vosos*

Ma2pms→Posesivo, determinante/non determinante, 2ª, plural, masculino, singular: *vosos*

Ma2sfp→Posesivo, determinante/non determinante, 2ª, singular, feminino, plural: *túas*

Ma2sfs→Posesivo, determinante/non determinante, 2ª, singular, feminino, singular: *túa*

Ma2smp→Posesivo, determinante/non determinante, 2ª, singular, masculino, plural: *teus*

Ma2sms→Posesivo, determinante/non determinante, 2ª, singular, masculino, singular: *teu*

Ma3afp→Posesivo, determinante/non determinante, 3ª, singular/plural, feminino, plural: *súas*

Ma3afs→Posesivo, determinante/non determinante, 3ª, singular/plural, feminino, singular: *súa*

Ma3amp→Posesivo, determinante/non determinante, 3ª, singular/plural, masculino, plural: *seus*

Ma3ams→Posesivo, determinante/non determinante, 3ª, singular/plural, masculino, singular: *seu*

Md1sap→Posesivo, determinante, 1ª, singular, masculino/feminino, plural: *mis*

Md1sas→Posesivo, determinante, 1ª, singular, masculino/feminino, singular: *mi*

Md2sap→Posesivo, determinante, 2ª, singular, masculino/feminino, plural: *tus*

Md2sas→Posesivo, determinante, 2ª, singular, masculino/feminino, singular: *tu*

Md3aap→Posesivo, determinante, 3ª, singular/plural, masculino/feminino, plural: *sus*

Md3aas→Posesivo, determinante, 3ª, singular/plural, masculino/feminino, singular: *su*

NUMERAL (N) (9 etiquetas)

Ncaap→Numeral, cardinal, determinante/non determinante, masculino/feminino, plural: *tres, quatro...*

Ncafp→Numeral, cardinal, determinante/non determinante, feminino, plural: *dúas, trescentas...*

Ncafs→Numeral, cardinal, determinante/non determinante, feminino, singular: *unha*

Ncamp→Numeral, cardinal, determinante/non determinante, masculino, plural: *dous, douscentos...*

Ncams→Numeral, cardinal, determinante/non determinante, masculino, singular: *un*

Noafp→Numeral, ordinal, determinante/non determinante, feminino, plural: *primeiras, terceiras...*

Noafs→Numeral, ordinal, determinante/non determinante, feminino, singular: *primeira, terceira...*

Noamp→Numeral, ordinal, determinante/non determinante, masculino, plural: *primeiros, terceiros...*

Noams→Numeral, ordinal, determinante/non determinante, masculino, singular: *primeiro, terceiro...*

PRONOME (R) (13 etiquetas)

Rtr1ap→Pronome, tónico, nominativo/preposicional, 1ª, masculino/feminino, plural: *nós*

Rtr1fp→Pronome, tónico, nominativo/preposicional, 1ª, feminino, plural: *nosoutras*

Rtr1mp→Pronome, tónico, nominativo/preposicional, 1ª, masculino, plural: *nosoutros*

Rtr2ap→Pronome, tónico, nominativo/preposicional, 2ª, masculino/feminino, plural: *vós*

Rtr2as→Pronome, tónico, nominativo/preposicional, 2ª, masculino/feminino, singular: *ti*

Rtr2fp→Pronome, tónico, nominativo/preposicional, 2ª, feminino, plural: *vosoutras*

Rtr2mp→Pronome, tónico, nominativo/preposicional, 2ª, masculino, plural: *vosoutros*

Rtr3ap→Pronome, tónico, nominativo/preposicional, 3ª, masculino/feminino, plural: *vostedes*

Rtr3as→Pronome, tónico, nominativo/preposicional, 3ª, masculino/feminino, singular: *vostede*

Rtr3fp→Pronome, tónico, nominativo/preposicional, 3ª, feminino, plural: *elas*

Rtr3fs→Pronome, tónico, nominativo/preposicional, 3ª, feminino, singular: *ela*

Rtr3mp→Pronome, tónico, nominativo/preposicional, 3ª, masculino, plural: *eles*

Rtr3ms→Pronome, tónico, nominativo/preposicional, 3ª, masculino, singular: *el*

RELATIVO (T) (4 etiquetas)

Tafp→Relativo, determinante/non determinante, feminino, plural: *cantas*

Tafs→Relativo, determinante/non determinante, feminino, singular: *canta*

Tamp→Relativo, determinante/non determinante, masculino, plural: *cantos*

Tams→Relativo, determinante/non determinante, masculino, singular: *canto*

VERBO (V) (3 etiquetas)

V0f10s→Verbo, tempo non aplica, infinitivo, 1ª, xénero non aplica, singular: *cantar, coser, partir...*

V0f30s→Verbo, tempo non aplica, infinitivo, 3ª, xénero non aplica, singular: *cantar, coser, partir...*

V0fa0s→Verbo, tempo non aplica, infinitivo, 1ª/3ª, xénero non aplica, singular: *cantar, coser, partir...*

ADVERBIO (W) (1 etiqueta)

Wa→Adverbio, nuclear/modificador: *moito, terriblemente...*

Apéndice E

Secuencias válidas de pronomes enclíticos en XIADA

A táboa que presentamos contén, tal como indica o encabezamento do apéndice, a relación de secuencias de pronomes enclíticos válidas en galego; está ordenada alfabeticamente en primeiro termo pola *secuencia* e en segundo pola *lonxitude* da cadea; os valores numéricos da segunda columna indican o número de clíticos dos que consta a secuencia. Para a súa elaboración partimos de Álvarez, Regueira e Monteagudo (1986) e de Álvarez (1994).

<i>Secuencia</i>	<i>Lonxitude</i>
a	1
as	1
cha	2
chas	2
che	1
chella	3
chellas	3
chelle	2
chelleme	3
chellenola	4
chelleno-la	4
chellenolas	4
chelleno-las	4
chellenolo	4
chelleno-lo	4
chellenolos	4
chelleno-los	4
chellenos	3
chellevola	4
chellevo-la	4
chellevolas	4
chellevo-las	4
chellevolo	4
chellevo-lo	4
chellevolos	4
chellevo-los	4
chellevos	3
chello	3

<i>Secuencia</i>	<i>Lonxitude</i>
chellos	3
chema	3
chemas	3
cheme	2
chemelle	3
chemelles	3
chemevo-la	4
chemevo-las	4
chemevo-lo	4
chemevo-los	4
chemevos	3
chemo	3
chemos	3
chenola	3
cheno-la	3
chenolas	3
cheno-las	3
chenolo	3
cheno-lo	3
chenolos	3
cheno-los	3
chenos	2
chevola	3
chevo-la	3
chevolas	3
chevo-las	3
chevolo	3
chevo-lo	3

<i>Secuencia</i>	<i>Lonxitude</i>
chevolos	3
chevo-los	3
chevos	2
cho	2
chos	2
la	1
-la	1
las	1
-las	1
lla	2
llas	2
lle	1
llela	2
lle-la	2
llelas	2
lle-las	2
llelo	2
lle-lo	2
llelos	2
lle-los	2
llema	3
llemas	3
lleme	2
llemo	3
llemos	3
llenola	3
lleno-la	3
llenolas	3
lleno-las	3
llenolo	3
lleno-lo	3
llenolos	3
lleno-los	3
lleos	2
lles	1
llevola	3
llevo-la	3
llevolas	3
llevo-las	3
llevolo	3
llevo-lo	3
llevolos	3
llevo-los	3

<i>Secuencia</i>	<i>Lonxitude</i>
llevos	2
llo	2
llos	2
lo	1
-lo	1
los	1
-los	1
ma	2
mas	2
me	1
mella	3
mellas	3
melle	2
mellela	3
melle-la	3
mellelas	3
melle-las	3
mellelo	3
melle-lo	3
mellelos	3
melle-los	3
melles	2
mello	3
mellos	3
mevola	3
mevo-la	3
mevolas	3
mevo-las	3
mevolo	3
mevo-lo	3
mevolos	3
mevo-los	3
mevos	2
mo	2
mos	2
na	1
nas	1
no	1
nola	2
no-la	2
nolas	2
no-las	2
nolo	2

<i>Secuencia</i>	<i>Lonxitude</i>
no-lo	2
nolos	2
no-los	2
nos	1
noslles	2
nosvos	2
o	1
os	1
se	1
seche	2
sechelle	3
sechelles	3
sechemevos	4
sechenos	3
sechevos	3
selle	2
selle-la	3
selle-las	3
selle-lo	3
selle-los	3
selles	2
seme	2
seno-la	3
seno-las	3

<i>Secuencia</i>	<i>Lonxitude</i>
seno-lo	3
seno-los	3
senos	2
sevo-la	3
sevo-las	3
sevo-lo	3
sevo-los	3
sevos	2
te	1
telle	2
telles	2
teme	2
temelle	3
tenos	2
vola	2
vo-la	2
volas	2
vo-las	2
volos	2
vo-lo	2
vo-los	2
vos	1

Apéndice F

Ocorrencias dos verbos *informar* e *instar* na súa construción con complemento de réxime no CORGA

Recollemos neste apéndice todas as aparicións documentadas na versión 1.5 do CORGA das construcións con complemento de réxime ou suplemento dos verbos *informar* e *instar* con pronome de terceira ou co de segunda singular, persoas nas que a oposición de caso, acusativo fronte a dativo, se manifesta na expresión. Incluímos, así mesmo, a casuística completa destas construcións cando a pesar de non existir a pronominalización si está presente o complemento de persoa, ben con preposición *a* ben sen preposición, excepto con *instar* na prensa do CORGA onde por ser excesivos os casos só recollemos unha mostra por ano e xornal. Así pois, prescindimos das secuencias en que aparecen os pronomes de primeira singular ou plural e segunda plural por producirse nesas persoas sincretismo formal. Tamén eliminamos desta listaxe os exemplos ambiguos co pronome *nos* que tanto poden corresponder ao alomorfo de terceira acusativo masculino plural como ao pronome acusativo/dativo de primeira plural:

Despois de servirnos dúas racións do peor whisky que se podía atopar en Lisboa, o camareiro da tasca informounos de que a Fundación Gisela Antunes era un lugar tranquilo. [CORGA: A. Malvar, *Un home que xaceu aquí*, Sotelo Blanco, 1995]

O afable cabaleiro, estudante do primeiro curso de medicina na Universidade de Salamanca, informounos do descubrimento de novas terras africanas que os navegantes portugueses fixeran nos pasados anos. [CORGA: X. C. Fernández Caínzos, *Os espidos pés do destino*, Bahía, 2000]

Por suposto, as ocorrencias de *nos* nas que o contexto aclara que se corresponden coa terceira persoa, e non coa primeira plural, son contabilizados na epígrafe correspondente. Por exemplo:

Así mesmo, destacou que "existe unha significativa e continua ameaza por parte dos grupos paramilitares que non aceptan o proceso de paz" e instounos para que se desarmen o antes posible. [CORGA: CG2001-10-25/26]

A listaxe de ambas as estruturas está organizada en cinco pequenos bloques:

- subcorpus de adestramento xornalístico

- subcorpus de adestramento de ficción
- revistas do CORGA
- xornais do CORGA
- ficción do CORGA

Ademais, no interior de cada bloque, separamos os exemplos en función de se a pronominalización se dá en forma acusativa ou en forma dativa e, en caso de non existir esta, clasificamos a presenza do complemento de persoa segundo presente ou non a preposición *a*.

1) INFORMAR (A) ALGUÉN DE / SOBRE ALGO

a) no corpus de adestramento xornalístico

Pronominalización en forma acusativa:

Os directivos de Endesa e o conselleiro delegado de Iberdrola, Javier Herrero, reuníronse onte pola mañá cos conselleiros de Economía e Industria da Xunta, Xosé Antonio Orza e Juan Rodríguez Yuste, para informalos sobre as consecuencias en Galicia do proceso de fusión. [CORGAetq: CG2000-12-02/10]

Noutra orde de cousas, o ministro de Economía, Rodrigo Rato, admitiu onte que a ex presidenta da CNMV, Pilar Valiente, deba informalo das xestións do ex secretario de Estado de Facenda Enrique Giménez-Reyna sobre Gescatera, e que esa falta de información foi unha das razóns polas que aceptou a súa dimisión. [CORGAetq: CG2001-10-04/33]

En declaracións a Antena 3, e en resposta ás preguntas sobre se Valiente deba informalo das xestións que Giménez-Reyna estaba facendo ante a presidenta da CNMV en relación con Gescatera e se esa foi unha das razóns polas que aceptou a súa dimisión, o ministro respondeu con dous concisos "si". [CORGAetq: CG2001-10-04/33]

O tribunal que xulga o 11-M impuxo antonte unha multa de 1.000 euros e ordenou deducir testemuño contra Díaz de Mera por un delito de desobediencia grave, despois de que este se negase a revelar o nome da persoa que o informou da existencia dun suposto informe sobre vinculacións entre ETA e os atentados. [CORGAetq: GH2007-03-30/55]

Pronominalización en forma dativa:

Aínda que as reivindicacións duns e doutros son diferentes a suma de ambas as protestas afectou a numerosos pacientes, que onte se queixaban fundamentalmente de que ninguén lles informou da xornada de paro. [CORGAetq: GH2007-03-30/8]

Sen pronominalizar, complemento sen preposición:

Pola súa parte, Francisco Vázquez explicou que **informou** os representantes do comité de empresa **de** cáál é a situación das negociacións con posibles investidores e referiuse á necesidade de que "ninguén baixe a garda". [CORGAetq: CG2001-10-10/3]

O Executivo comunitario **informou a comisión de Presupostos do PE** **de que** a finais deste mes proporá utilizar para financia-lo dito plan o denominado "instrumento de flexibilidade" previsto no presuposto comunitario para gastos excepcionais.[CORGAetq: CG2001-10-18/1]

Sen pronominalizar, complemento con preposición:

Fontes dos convocantes da manifestación (entre outros, o Foro Social de Madrid, PSOE, IU, UGT e CCOO) **informaron a Efe** **de que** na cabeceira da marcha situaranse esas persoas ataviadas con fundas laranxas e que portarán unha pancarta co lema Pola paz, non á guerra. [CORGAetq: GH2007-03-17/61]

b) no corpus de adestramento de ficción

Fronte a toda a variación do corpus de adestramento xornalístico, no corpus seleccionado para o adestramento da ficción, na estrutura “informar alguén de algo” só se documenta un caso en cuxa expresión pronominalizada aparece a forma de acusativo:

Presentárono ao gran público (despois de informalo, polo miúdo, dos avances da civilización) nunha rolda de prensa que celebraron, por deferencia co entrevistado, no Castro de Baroña, a media tarde dun caloroso e concorrido domingo de verán. [CORGAetq: o_xene_da_chuvia/29]

Con todo, no haber desta acepción, debemos sumar outros dous casos nos que, malia non estar o complemento de réxime explícito, si se deben contabilizar por aparecer pronominalizado o complemento de persoa en forma de acusativo, feito que nos asegura estar ante a estrutura “informar (a) alguén de algo”:

Pedireille me concedan xa pensión viuvez. Informareite. Deica outra." [CORGAetq: de_corpo_enteiro/5]

O señor Quintela informaría; nada de preguntarlle ao axudante, que nunca sabía nada. [CORGAetq: vidas_exemplares_18]

c) nas revistas do CORGA

Pronominalización en forma acusativa:

Este alto mandatario informouno no encontro das intencións da COTOPV de facer alí "un pequeno encoro". [CORGA: GI1996-04/38]

Comunícase co Pool para proporcionarlle recursos e informalo de modificacións nestes. [CORGA: *DX2001-01/11*]

Nestes casos hai que se pór en contacto decontado coa compañía distribuidora para informala do acontecido. [CORGA: *C2006-10/45*]

Os medios non só establecen vínculos de relación entre individuos, grupos e sociedades informándoos de (certas) actividades e acontecementos. [CORGA: *TN2006-05/11*]

Pronominalización en forma dativa:

O obxectivo destas entrevistas foi **informarles a todos eles da** reclamación oficial presentada e manifestarlle-lo punto de vista de Teatro do Noroeste reflectido nun escrito que a todos se entregou e que é o seguinte: [CORGA: *TDN1997-02/2*]

Esixa que lle informen do que comeron e/ou merendaron a cotío; [CORGA: *C2005-09/2*]

d) nos xornais do CORGA

Pronominalización en forma acusativa:

Os pais pretenden que os informen sobre os casos epidemiolóxicos que se rexistran na contorna, e que "se nos facilite o rexistro da vacina subministrada ós nenos", así como as medidas a adoptar de contado na comarca. [CORGA: *CG1995-02-11/37*]

Xosé Cuiña, pola súa parte, anunciou que se poñería en contacto con Romay Beccaría e con Manuel Fraga, para informalos da postura mantida por estes concelleiros. [CORGA: *CG1995-02-15/20*]

Barrionuevo informouno de que se ía reforza-la vixilancia na zona con efectivos da Garda Civil, pero cando horas despois desta conversación entre Barrionuevo e Fraga se produciu o atentado, o reforzo aínda non chegara. [CORGA: *CG1995-02-26/14*]

Os arquitectos que elaboran o estudio entrevistáronse con técnicos municipais e da Confederación Hidrográfica do Norte para informalos respecto das características técnicas da obra. [CORGA: *CG1996-02-04/28*]

Xa en Cataluña temos que informalos da celebración, o día 23 de febreiro, da Gran Noite de Galicia en Barcelona. [CORGA: *CG1996-02-04/55*]

Fernández Toxo, que asistiu na Coruña a unha reunión de delegados sindicais para informalos do proceso da negociación colectiva, asegurou que nestes momentos o Goberno "non ten comprador nin capacidade para vender esta empresa antes das eleccións". [CORGA: *CG1996-02-09/41*]

Pola súa parte, o primeiro ministro irlandés, John Bruton, reunírase co presidente Clinton en marzo para informalo sobre o proceso de paz. [CORGA: *CG1996-02-15/63*]

Xa os informaremos. [CORGA: CG1996-02-18/36]

"Estou agora en condicións de informalo de que o comandante da Ifor (a forza internacional de paz en Bosnia) asegura que as forzas serbobosníacas se retiraron das zonas de separación establecidas polo acordo de paz (de Dayton)", di o escrito de Solana. [CORGA: CG1996-02-27/83]

Reprochoulle a Belloch que non a recibira para informala do rapto [CORGA: CG1996-02-28/88]

A Asemblea para Ceibar os Presos-as Políticos Galegos (ACPG), manifestouse onte ante o Valedor do Pobo, co gallo de informalo da gravísima situación que están a atravesar os seus familiares e amigos. [CORGA: CG1996-02-29/55]

Ademais, informouno sobre o turismo rural e, concretamente, sobre a Casa de Láncara, onde as dependencias nas que se aloxou Castro durante o 92 foron bautizadas como 'Cuarto do Comandante'. [CORGA: CG1999-05-12/32]

En Washington mantivo contactos con altos funcionarios do Banco Mundial (BM) e do Banco Interamericano de Desenvolvemento (BID) para informalos da celebración o vindeiro ano do Foro Bolívar en Galicia e solicita-lo seu apoio á súa organización. [CORGA: CG1999-05-21/25]

Analiza-los procedementos técnicos necesarios e realiza-la hixiene persoal dun paciente ou cliente en función do seu estado, idade e/ou situación, informándoo da técnica a realizar e respectando a súa intimidade. [CORGA: DOG1999-03-22/1]

Nesta semana celebráranse nas tres universidades galegas diversas actividades para anima-los estudantes a crea-las súas propias empresas e a informalos sobre as axudas da consellería para fomenta-lo emprego autónomo, das que 13.000 persoas se beneficiaron en 1998 e 1999. [CORGA: CG2000-11-14/84]

María Ángeles Villar manifestou que o representante de Scotland Yard, Kevin Power, púxose en contacto con eles para informalos da sesión desta semana, que non estaba prevista en principio, e agregou que noutras ocasións se celebran igual sen que eles reciban aviso ningún. [CORGA: CG2000-11-27/15]

Os directivos de Endesa e o conselleiro delegado de Iberdrola, Javier Herrero, reuníronse onte pola mañá cos conselleiros de Economía e Industria da Xunta, Xosé Antonio Orza e Juan Rodríguez Yuste, para informalos sobre as consecuencias en Galicia do proceso de fusión. [CORGA: CG2000-12-02/10]

Nun novo encontro o 17 de abril, Aznar informouno das conversas con CiU e PNV, así como das propostas feitas no financiamento autonómico. [CORGA: CG2000-12-18/4]

O presidente de EEUU, Bill Clinton, que reiterou a súa disposición a axudalos a alcanza-lo

acordo, falou onte por teléfono co seu colega exipcio, Hosni Mubarak, para informalo de cómo evolucionan as conversacións que as partes teñen agora na base aérea de Bolling, preto de Wáshington. [CORGA: CG2000-12-18/4]

O delegado do Goberno, Arsenio Fernández de Mesa, acompañado polo conselleiro de Presidencia, Xaime Pita, convocou na sede coruñesa da Delegación do Goberno a alcaldes de toda Galicia para informalos sobre o novo prazo de solicitude de axudas estatais e despexa-las dúbidas que puidesen xurdir ó respecto sobre elas. [CORGA: CG2000-12-27/2]

En declaracións a Antena 3, e en resposta ás preguntas sobre se Valiente debera informalo das xestións que Giménez-Reyna estaba facendo ante a presidenta da CNMV en relación con Gescartera e se esa foi unha das razóns polas que aceptou a súa dimisión, o ministro respondeu con dous concisos "si". [CORGA: CG2001-10-04/33]

Fontes do Sindicato Español de Pilotos de Liñas Aéreas (SEPLA) confirmaron que a sección sindical en Iberia manterá o luns, día 15, unha reunión coa compañía para informalos das medidas que a aeroliña vai tomar. [CORGA: CG2001-10-14/11]

Nesta liña, o director de SIMO TCI explicou que, como cada ano, se dedicará un día ás pemes co obxectivo de informalas do "ratio de eficacia que van gañar, así como do aforro de tempo e da entrada na era dixital" mediante a aplicación das Novas Tecnoloxías á súa actividade cotiá. [CORGA: CG2001-10-27/48]

Sharon reuniuse onte cos membros do 'minigabinete' de seguridade nacional _composto polos líderes de cada partido dentro da súa coalición de goberno_ para prepara-la visita a Wáshington e informalos da súa iniciativa de paz en termos xerais. [CORGA: CG2002-05-04/39]

A Audiencia considerou que non había motivos para que o prezo tivera incrementos tan substanciais e que os tres enganaran ó consello da Corporación para que aprobaran as operacións, ó non informalos do seu interese persoal e o beneficio que lles ía reportar. [CORGA: CG2002-05-10/11]

Nunha carta ó presidente Bush, o primeiro ministro indio informouno de que a súa capacidade para aguantar acababa, lamentándose de que o presidente paquistaní, Pervez Musharraf, non fixera nada para impedi-la entrada na Caxemira india dos rebeldes islámicos. [CORGA: CG2002-05-25/48]

Doutra, porén, a administración galega acusou á Generalitat de cometer unha "falta de cortesía" coa Xunta por non informala do envío dos tres barcos. [CORGA: ANT2003-01-16/16]

Nesa altura eu, pasados os primeiros días, pónome en contacto co presidente e, dunha maneira abondo elocuente, doulle pé para que chame a oposición, informala de como está a situación e facer fronte conxuntamente ao gravísimo problema. [CORGA: ANT2003-02-20/31]

Santiso Miramontes fixo este anuncio no transcurso dunha xuntanza cos responsables dos

consellos reguladores das denominacións de orixe e das indicacións xeográficas protexidas dos produtos galegos, para informalos das novidades recollidas na Lei de promoción e defensa da calidade alimentaria galega, aprobada polo Parlamento e de próxima publicación no DOG. [CORGA: GH2005-02-18/10]

Ao volver o soberano e informalo do activo proselitismo do santo, montou en cólera. [CORGA: GH2005-02-18/10]

Sen dúbida un pracer para os sentidos do que podes informarte máis, a través da súa web www.hotelpousadadelcastillo.com. [CORGA: GH2005-02-25/60]

Precisamente, "en canto ó pai que di isto, a infección confirmouse o sábado pola noite e non deu tempo a informalo desde o hospital, porque o domingo estaba en todos os xornais". [CORGA: GH2007-03-06/41]

Neste sentido, apuntou que nos seus encontros con empresarios no exterior informáraos dos apoios "de promoción económica e capital risco" cos que contan, a través de organismos da Administración autonómica, como poden ser o Igape e Xesgalicia. [CORGA: GH2007-03-24/18]

Por isto falamos coa comunidade internacional e pedimos unha intervención directa no conflito, e por isto tamén tivemos unha reunión dos nosos embaixadores coa Unión Europea para informalos, dunha maneira moi clara sobre o que ocorre en Palestina. [CORGA: GH2008-08-06/77]

Desta forma, "o departamento de Asistencia Xurídica, no momento de detectar a publicación dunha sanción, poríase en contacto co socio informándoo dos feitos". [CORGA: GH2008-08-06/22]

Pronominalización en forma dativa:

Prega entón que lle informen dun lugar máis afastado ata onde poida continuar o seu sacrificio de peregrinaxe. [CORGA: CG2000-11-26/52]

En van pretenderon os veciños afectados que lles informasen sobre o ritmo e cantidade de auga a ceibar: [CORGA: CG2000-12-17/23]

Sen embargo, Pérez non puido evitar un "¡hai que horror!" ao informarlle do protagonismo da clase política. [CORGA: ANT2002-09-26/19]

Entón, o presidente da sala dirixiuse ao letrado para informarlle de que a defensa recoñecera a comprensión do galego, polo que o instou a continuar expresándose no noso idioma. [CORGA: ANT2002-11-07/2]

E experimentei a súa temperamental enerxía cando se alporizou connigo, ao informarlle de que a Torre de Goiáns, en Boiro, fora mercada polo Marqués de Revilla, Joaquín Otero

Goyanes, voltando así a ser posesión da liñaxe epónima. [CORGA: ANT2003-01-30/87]

Minutos antes de lanzar Quintana este envite a Zapatero, o alcalde do BNG en Pontevedra, Miguel Fernández Lores, deu conta do dossier que remitirá ó presidente da Xunta, Manuel Fraga, informándolle das leis vixentes que conculcan os proxectos de Ence en Lourizán. [CORGA: GH2005-02-12/9]

Pronominalización en forma dativa, con frase preposicional explícita ou marca de impersonalidade:

Os tres membros do Consello de Administración, que rectificaron partes das súas anteriores declaracións ante a xuíza Teresa Palacios ante preguntas das acusacións, dixeron que se lles informara dos contratos antes do primeiro Consello de Administración do banco fusionado e que os viron entre os documentos que se acompañan a estas reunións. [CORGA: GH2005-02-08/50]

O concelleiro de Seguridade Cidadá de Villajoyosa, Pedro Alemany, informoulle a Efe de que a citada residencia só é ocupada por persoal da entidade bancaria e familiares entre os meses de maio a outubro. [CORGA: GH2005-02-28/29]

"O venres ás 11.30 xa se lle informara a todos os portavoces e ó presidente da Comisión de que se retirara porque non tiña sentido volver a formulala". [CORGA: GH2007-03-20/15]

Sen pronominalizar, complemento sen preposición:

O departamento de Servicios Sociais do Concello de Ribeira está procedendo a informa-los anciáns sobre a concesión gratuíta que, para gafas e audífonos, promove a Consellería de Sanidade e Servicios Sociais dentro da súa campaña Programa para a mellora visual e auditiva das persoas maiores. [CORGA: CG1995-02-07/54]

Os comerciantes locais foron convocados polo BNG a unha reunión o próximo martes, día 14, para estudar co empresariado local as alegacións a presentar e para informa-los veciños das posibilidades do recurso. [CORGA: CG1995-02-11/40]

En Hong Kong, un funcionario do Banco de Inglaterra chegou para informa-las autoridades da bancarrota de Barings e a situación resultante ó non poderse atopar unha solución ó problema. [CORGA: CG1995-02-28/71]

Pretende Ravachol informa-los pontevedreses, que ata este lugar se acheguen, sobre os detalles da remodelación da praza de touros de San Roque, para que logo non xurdan disquisicións acerca da funcionalidade do proxecto. [CORGA: CG1995-02-28/74]

Estes lemas, como se recordará foron utilizados, respectivamente, na campaña organizada pola Joventut Nacionalista de Catalunya (JNC) durante os Xogos Olímpicos e na organizada para informa-los turistas sobre Cataluña. [CORGA: CG1996-02-17/75]

Así mesmo, as comunidades autónomas e as entidades locais deberán informa-los consumidores sobre os perigos da eliminación incontrolada e sobre a eliminación correcta dos aparellos que incorporen este tipo de materias. [CORGA: CG1996-02-27/145]

Segundo a súa opinión, é preciso informa-lo consumidor de que, sen necesidade de realizar investimentos, é posible reduci-lo seu consumo facendo un bo uso da enerxía sen reduci-lo seu nivel de confort. [CORGA: CG1996-02-27/147]

Así, en abril iniciouse unha campaña nos medios de comunicación con anuncios que informan os máis novos da importancia de actuar correctamente durante a conducción. [CORGA: CG1999-05-02/30]

O acto, que se enmarca na campaña nacional 'Despois... Non hai marcha atrás', ten como obxectivo informa-los mozos sobre os métodos anticonceptivos ante o alarmante aumento de embarazos de adolescentes e de infectados pola sida. [CORGA: CG2000-11-02/41]

Por outra parte, Mayor Oreja asistirá hoxe á reunión plenaria que celebra o Grupo Popular no Congreso para informa-los parlamentarios do PP da política antiterrorista do Goberno, explicáronlle a Europa Press en fontes do PP. [CORGA: CG2000-11-21/46]

En calquera caso, deixou claro que, nestes encontros bilaterais, Traballo non retirou o seu documento, e engadiu que a semana que vén, o Goberno informará os axentes sociais de cal é a súa posición e da decisión que adoptará en relación con esta reforma. [CORGA: CG2001-10-12/29]

Ademais, pediulles ós socialistas que informen os cidadáns sobre as súas propostas Verbo disto. [CORGA: GH2005-02-08/12]

Sen pronominalizar, complemento con preposición:

Este órgano colexiado informará ás asociacións xuvenís puntualmente dos acordos acadados e recibirá das citadas asociacións suxestións para a súa incorporación ós programas dirixidos á mocidade. [CORGA: CG2000-11-05/36]

Ugaz asumiu o venres as súas funcións despois de que o Goberno de Suíza informara ó de Lima de que Montesinos ten alomenos tres contas noutros tantos bancos dese país europeo con 48 millóns de dólares en depósitos de orixe sospeitosa. [CORGA: CG2000-11-05/36]

Se ben catro anos antes xa informara eu a Filgueira Valverde da existencia daquel poema case inédito de Otero Pedrayo. [CORGA: CG2000-11-07/55]

"Onishi reuniu os pilotos en fronte da casa de Santos asegurándolles que a súa intervención cambiaría o curso da guerra e que informaría persoalmente diso ó emperador", manifesta Daniel Dizon. [CORGA: CG2000-11-13/63]

Os seus socios informarán aos cidadáns dos dereitos e de ser preciso denunciarán a Francisco

Vázquez por desobediencia. [CORGA: ANT2003-02-06/54]

Tamén informará a estes grupos dos recentes contactos mantidos entre representantes da ANP e Israel. [CORGA: GH2005-02-12/38]

"España informará á Comisión, nun prazo de dous meses a partir da data de notificación da presente decisión, das medidas previstas e xa adoptadas en cumprimento desta. [CORGA: ANT2007-03-15/12]

Pero, a maior abundamento, os editores da NLR, na presentación que fan dese documento, informan aos seus lectores de que a tal entrevista fora publicada meses antes no xornal Haaretz baixo estoutro rótulo: "Supervivencia dos máis aptos" _enunciado que, inevitábelmente, fainos evocar o darwinismo social racista e, por pasiva, o designio nazi da[texto omitido] [CORGA: GH2007-03-18/48]

e) na ficción do CORGA

Pronominalización en forma de acusativo:

[CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Acontecera, que ó començar a disputa entre Carlos i o periodista colombiano, o aragonés chamóu por teléfono a éste e informóuno de que Montemar era Carlos Aranda Torres;

Agora mesmo vou a infórmalo desta falcatuada.

Informéino dos receos de Firmino, i el tranquilizóume coa espricación que me deu.

Dínlle o meu nome e informéino de que traballaba en sociedade co comandante Lugo.

Cabo dunha hora volvéu túa nai _sigue decido Xulia_ i o Alcázar, ó vela, informóuna de que non estabas eiquí;

[CORGA: V. Fernández Freixanes, *O triángulo inscrito na circunferencia*, Galaxia, 1982]

Paréceme honesto, lector, informarte dende o principio de a ónde nos pode levar todo isto se alguén non o remedia, en qué aventura te estás a meter se segues a bordo.

[CORGA: C. Gonsar, *Desfeita. (Semirreportaxe)*, Xerais, 1983]

Informámolo do lume.

[CORGA: M^a X. Queizán, *Amantia*, Xerais, 1984]

A criada informouna que a súa avoa estaba agardánda no tablinium cun señor.

[CORGA: C. G. Reigosa, *Crime en Compostela*, Xerais, 1984]

_Ben, da vida que fan durante o día xa te informarán por aí.

[CORGA: X. R. Pena, *O reverso do espello*, Xerais, 1984]

_Descanse _condescendeu gustoso_, descanse, D. Pedro, e disculpe vostede a miña premura en informalo acerca de todo en tan breve espacio.

Pero permítame informalo de que sen dúbida ten vostede grandes dotes de convicción.

[CORGA: R. Raña Lama, *Nun cuarto para agardar*, Edicións do Castro, 1985]

Desde que, no día anterior, o informaran de que o seu pai aparecera morto, en estado de putrefacción, na illa de Brooklyn, envenenado por si mesmo, a visión da morte do pai era cíclica e infinita, acontecía como un dardo asiduo aloxado no adentro do seu ser.

[CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]

_Hai tres feridos graves _informouno Pedro, un dos tres Pedros que os seguiran_.

Pedro informouno do que acontecerá.

[CORGA: Q. Cadaval e X. Lama, *Memoria de Antígona*, Xunta de Galicia, 1988]

Antígona: Infórmante ben.

[CORGA: X. M. Martínez Oca, *As florestas do Mañuema*, Xerais, 1988]

O guerreiro que apresaran informáraos de que o sacrificio dos prisioneiros produciríase ó abrir o sol e sería a propia Momoa a encargada de executalo, ofrecendo ó deus o sangue das vítimas.

No Mañuema terían que localizar ó doutor Livingstone antes de que se volvea perder e informalo da expedición que chegaba na súa axuda.

[CORGA: F. X. Fernández Naval, *O bosque das antas*, Xerais, 1990]

Fíxoo un xastre cando o xuicio informouno Amador.

[CORGA: M. Forcadela, *Paisaxe con muller e barco*, Xerais, 1990]

O doutor Frazer informouno máis tarde sobre as características da súa doenza que non dubidou en calificar como moi grave ao tempo que sinalou as súas especiais vinculacións coa guerra pasada e a súa vida familiar.

[CORGA: X. Alcalá, *Latitude austral*, Galaxia, 1991]

Don Pedro deulle a calada por resposta a respecto de bondades ou ruindades, e informouno de que a súa misión sería conducir un suxeito nun coche rápido moi lonxe, con diñeiro na bolsa abondo como para ir polo camiño mudando cabalos co mellor que houbera en postas e estancias do percorrido.

[CORGA: Manuel María, *Auto do Castromil ou a revolución dos baúles*, Transportes Castromil, 1992]

Baúl innominado: E non podo informarte.

[CORGA: X. Alvilares, *Memorial de agravios*, Galaxia, 1992]

Respecto ó clero que nos elixira, nós sentiámonos cun dobre deber: informalo de canto sucedera, máxime cando os perdedores de Lugo promoveran a acusación da ilexitimidade da nosa elección, e de comprometelo activamente nos cambios que a Asemblea Nacional propuxera.

[CORGA: B. Iglesias Araúxo, *Vento de seda*, Xerais, 1992]

O taberneiro informouno de que a irmá estivera a preguntar por el e que andaba feita unha madanela.

[CORGA: A. Álvarez Cáccamo, *Xente de mala morte*, Galaxia, 1993]

Teremos tempo de informarte.

[CORGA: M. Riveiro Loureiro, *Corpo canso*, Ir Indo, 1993]

Faloulles longamente valéndose dun megáfono, informándoos polo miúdo do moito que el se desvive, moitísimo, polo ben do pobo.

[CORGA: M. Forcadela, *Fóra de xogo*, Edicións do Cumio, 1993]

Mire son un empregado do estadio de Balaidos e quero informalo de que hoxe deixou vostede unha carteira coa súa documentación na bancada de tribuna.

[CORGA: R. Fonte, *Os leopardos da lúa. Unha traxedia particular*, Galaxia, 1993]

Prometeulle que non se volvería meter en líos, e o seu pai informouno de que andaba todo moi revoltado no arsenal, e que xa sabía el que cando os militares _neses momentos o seu pai púñase á marxe da súa condición de sarxento_ andaban nerviosos, todo podía ocorrer.

[CORGA: M. Aleixandre, *Tránsito dos gramáticos*, Xerais, 1993]

_As miñas posibilidades de traballo aquí, que son escasas _respondeu ela_, non dependen de Xosé Manuel, e non teño por que darlle explicacións, só informalo dunha decisión tomada.

[CORGA: E. R. Ruibal, *De corpo enteiro*, Xerais, 1994]

Informareite.

[CORGA: X. Cermeño, *Ciencia de facer as camas*, Xerais, 1994]

gastei os cartos que tiña en coller un taxi ata a miña casa e chamar por teléfono os amigos para informalos.

[CORGA: F. Alonso, *Cemiterio de elefantes*, Xerais, 1995]

Na farmacia informáranos de que en Pizarro había unha clínica dental duns arxentinos que estaba aberta toda a noite.

[CORGA: M. Aleixandre, *Lobos nas illas*, Xerais, 1996]

A expresión do meu debeu traizoarme porque engadiu, talvez o informaron mal.

[CORGA: A. Fernández Paz, *Avenida do Parque, 17*, SM, 1996]

Bea informouna tamén sobre os profesores, ela xa os coñecía polas referencias que lle deran algúns compañeiros que ían nos cursos máis adiantados.

[CORGA: X. Alcalá, *Código morse*, Xerais, 1996]

Do feito de preguntar deducín que non daba por suposto o normal nunha familia: xuntarse; e, para aliviarlle a conciencia, informeino de que tiña noiva en Lugo e tal vez fose cear coa xente dela.

E noso pai informouno, contundente:

[CORGA: X. Cid Cabido, *Oralmente pola boca*, Xerais, 1997]

O conductor presentouse como Alex (Martina imaxinou que Alesandre, porque non tiña aspecto nin acento estranxeiro) e informouna de que se dirixía a unha vila, de nome descoñecido para ela, a uns cincuenta quilómetros ó sur doutra cidade que Martina tampouco pisara nunca, en resumidas contas, case a mil cincocentos quilómetros de onde se atopaban.

[CORGA: X. M. Valcárcel, *Os ollos da sentinela*, Galaxia, 1997]

Oportunamente, e coa discreción que esixía un caso así, o pai informouna de que só podía facelo con un, pero dándolle a elixir.

E xa se sabe, os médicos non te informan con claridade, nin che din; todo se volve pasar duns a outros, facerte dar voltas e, total, para nada.

[CORGA: L. M. García Mañá, *O lume de Santo Antón*, Xerais, 1997]

Cando o cabo chegou ó mediodía xunta o seu xefe, informouno do fracaso da misión.

O sancristán da catedral atopouse con don Efrén na Praza do Ferro e informouno da mensaxe de Nemesio.

A recomendación de don Jovino foi que seguira a informalo.

Tamén lle faría chegar un escrito urxente ó bispo informándoo dos detalles das pescudas, demostrándolle que o poder civil tamén tiña recursos eficaces e, no fondo, gabándose de tomarlle a dianteira.

[CORGA: A. Fernández Paz, *O centro do labirinto*, Xerais, 1998]

Frida informouna dun xeito directo, sen reviravoltas.

[CORGA: X. C. Caneiro, *Un xogo de apócrifos*, Xerais, 1998]

Deseguida o informaron da imposibilidade desta suposición:

[CORGA: X. C. Caneiro, *A rosa de Borges*, Sotelo Blanco, 2000]

A miña cabeza preocupada naquela ocasión por ter que abandonar o refuxio onde o aburrimiento perseguido resultaba familiar e amable, un burato en medio do cotián e habitual, un pozo de paz, de calma, sen teléfono pitando, señor, debo informalo das últimas declaracións de tal e tal que argumenta de maneira vaga e imprecisa sobre as causas do deterioro das relacións con Marrocos por culpa do peixe e tal e tal.

[CORGA: X. C. Caneiro, *Ébora*, Xerais, 2000]

Fidel informouno da verdadeira personalidade da muller que solicitara os seus servicios, Matilde, que xogaba á brisca algunha vez no seu local.

Na Comisaría, onde todo está rexistrado, informárono correctamente.

[CORGA: R. Rei Núñez, *Expediente Artieda*, Xerais, 2000]

"Non che custaría informarte de quen foi."

¿Informaría Silverio do noso encontro?

[CORGA: X. Borrazás, *Na maleta*, Sotelo Blanco, 2000]

Se está aquí, informouno, é porque confiamos en vostede.

O caixeiro automático informouno de que o seu saldo limitábase agora á dezmilésima parte do que tiña na maleta, un chisco menos.

[CORGA: X. C. Fernández Cainzos, *Os espidos pés do destino*, Bahía, 2000]

A criada informouno de que María estaba prometida a un mozo que pertencía a unha familia acomodada de Burgos e ambos contraerían matrimonio no seguinte Nadal.

[CORGA: Jaureguizar, *As horas sucias*, Xerais, 2001]

"Antes diso, quería informalo de que é transpirable e de que repele o po e os ácaros."

[CORGA: A. Lopo, *Ganga*, Xerais, 2001]

Nunha leve paréntese da crónica, il informouno de que non consumía colesterol.

No despacho de servizo ó cliente, unha rapaza de cadencia narnexa informouno con desgana dos horarios

[CORGA: C. G. Reigosa, *Narcos*, Xerais, 2001]

Pablo recoñeceu a voz do seu irmán Vicente, que o informaba sen demoras introductorias.

informareino e nada máis, e cando non lle guste o que fago, considerareime despedido.

En cambio, cando non lle dicía nada, o home facía uns esforzos considerables por espertar o seu interese e informalo de cousas que puidesen merecer a súa atención.

Alexandre Míngallos, un mozo alto e forte, de vinteoito anos, de pelo rizo e ollada viva,

acababa de chegar de Bogotá, vía Miami, e informouno de que estaba solucionado o enlace (é dicir, o traslado) Colombia-Panamá.

Concluía que non podía facer nada mellor que chamar ó amo do fume de Grobas e informalo do que pasaba.

Pablo Mingallos, incómodo, informounas:

Colleuno e escoitou a voz inqueda e desacougada de seu irmán Vicente que o informaba dun movemento de gardas e policíaas por Valverde.

[CORGA: X. López López, *O caderno*, Sotelo Blanco, 2001]

Iker preguntou, e o noso escritorciño, sen lle dar máis transcendencia, informouno do seu proxecto:

[CORGA: X. Quiroga, *Atuado na braña*, Xerais, 2002]

Detrás da porta escoitou de novo a voz apurada de Lupe que o informaba de que era día de barbeiro.

A serventa informouno de que a señora saíra para asistir a unha reunión.

Ela informouno de que nese intre estaba ó teléfono o alcalde de Escairón, que xa chamara tres veces na última hora, Insistindo en que lle dese o número do teu móbil.

O Apache marchou informouno.

Colleulle o gabán e rapidamente informouno de que o Xeneral estaba reunido no salón principal.

Era a substituta de Marga que o informaba de que a súa nai, Vaia, ho, quería falar con el.

[CORGA: M. Fernández Mallo, *Cabilia*, Primerapersona, 2005]

O Presidente do Banco Bilbao contestara con certo interese, ao tempo que o informaba de que un directivo chamaríao axiña, tal como ocorreu.

[CORGA: R. X. Losada, *O xene da chuvia*, Edicións do Castro, 2005]

Presentárono ao gran público (despois de informalo, polo miúdo, dos avances da civilización) nunha rolda de prensa que celebraron, por deferencia co entrevistado, no Castro de Baroña, a media tarde dun caloroso e concorrido domingo de verán.

[CORGA: Jaureguizar, *Cabaret Voltaire*, Galaxia, 2005]

E mentres o asistente colocaba o servizo, Fermín Barnes, para quen as queixas de William Curst comezaban a ser unha comechón, informouno de que aquela cunca fora elaborada na firma de Ibáñez, un emprendedor dos Oscos asturianos.

Barnes informouno tamén de que xa posuía empresas ben rendibles en Ribadeo, que dera

cunhas canteiras de caolín ao pé das cales instalara obradoiros para producir esa cerámica que imitaba a inglesa.

E mentres o asistente lles servía as infusións, Fermín Barnes informouno de que aquela cunca era unha das moitas que se transportaran pola estrada de peaxe que creou Ibáñez en Sargadelos, na costa de Lugo.

E mentres o asistente lles ofrecía azucre para o café, Fermín Barnes informouno de que Ibáñez se decidira en 1788 a activar a súa creación de maior ambición ao pór en funcionamento a fábrica de ferros e aceiros de Sargadelos.

E mentres o asistente erguía o pano para ofrecerlles galletas, Fermín Barnes informouno de que aquela cunca en que ía beber o seu té procedía da empresa de cerámicas construída por un ilustrado que chegara da aldea de Ferreirela, preto de Santalla de Oscos, onde o seu pai era escribán de número.

O enxeñeiro informáralo por medio dun telegrama remitido dende París de que eran fuxidos e de que precisaban dun lar onde superar as dificultades do regreso.

K informouna entón de que esas comprobacións de documentación poderían levar días durante os que o requirido vivía nas oficinas gubernativas.

¿Informácheste?

Co tempo que levas connosco, non te informaches aínda da temperatura á que me gusta bañarme.

[CORGA: A. de la Cruz, *O descenso do derradeiro ocaso*, Sotelo Blanco, 2005]

_Xoán de Requeixo, o Resucitado da Carballeira _informouna Xema _.

[CORGA: U. Casal, *Vidas exemplares*, Xerais, 2006]

O señor Quintela informaríalo; nada de preguntarlle ao axudante, que nunca sabía nada.

[CORGA: X. Fuentes Castro, *O pescador de cana*, Toxosoutos, 2006]

Alguén os informara da miña pertenza a ese mundo, sempre singular e marxinal.

Mais antes, coido que sería bo informala dun incidente curioso.

O recepcionista informouna de que Phillip Verheeren había cinco semanas que abandonara o hotel sen deixar enderezo ningún.

Nunha das súas últimas cartas informábalo de que Nanette, a exquisita e querida Nanette, xa os deixara para sempre e que xacía nun antigo camposanto no que, de tarde en tarde, inhumaban os corpos de persoas que non dispuñan de sepultura no cemiterio moderno.

[CORGA: B. Graña, *Protoevanxeo do neto de Herodes*, Xerais, 2006]

Cando o informaron do matrimonio feito ás súas costas, ía un abafante día de calor e el acababa

de beber auga nun fermoso vaso de vidro, toda unha obra de arte que lle trouxeran de Alexandría, e entón sentiu de repente que xa non había dúbidas sobre a traizón do[texto omitido]

[CORGA: F. Castro, *Spam*, Galaxia, 2006]

velosojaneiro en realidade, viña correndo cara ao despacho de Gómez porque entendeu que había que llo dicir ao Xefe, informalo decontado antes de que a cousa fose a máis coma nos días pasados.

Pronominalización en forma dativa:

[CORGA: S. de Toro, *Tic tac*, A Nosa Terra, 1993]

María do Carmo Fernández Alvite, mestra de Preescolar no Colexio Público de Vilanova, diríxome a vostede para informarlle da situación na que se atopa a miña alumna M^a Vanesa Gerpe Añón, de 5 anos de idade.

Sen pronominalizar, complemento con preposición:

[CORGA: S. Santiago, *O silencio redimido. Historia dun home que pode ser outro*, Galaxia, 1976]

Entón non te demores, vaite, que eu informaréi á túa nai.

Na ficción do CORGA mesmo é posible atopar, xa non no mesmo autor senón incluso na mesma obra, a variación na escolla entre as dúas estruturas do verbo *informar*, o que provoca que apareza o pronome de dativo e o de acusativo:

- 1) informar (a) alguén de/sobre algo
- 2) informar algo a alguén

Vexamos os exemplos:

[CORGA: R. Raña, *O crime da rúa da moeda vella*, Galaxia, 1993]

Pensou que constituía un inmerecido, inenarrábel privilexio estar a falar con aquel xenio da música, do que coñecía a súa vida polos libros e revistas especializados, que o informaban de que gastara intimidade con Prokofiev, que coñecera persoalmente a Shostakovich, nomes lendarios que con só mencionalos o emocionaban até o aturdimiento

Teño que informar che, se o tes a ben, que a miña irmá Rosalía sofreu unha agresión esta tarde, que está no hospital, que me encontro moi afectado e que necesito estar só esta noite, sen que ninguén me preocupe con temas agora accesorios.

[CORGA: M. Lourenzo González, *Arqueofaxia*, Xerais, 1995]

O crego, daquela, pasou a informalos das actuacións do hospital en todos os seus pormenores.

Non existe ningún garaxe soterrado alí nin pola zona, segundo lle informan os viandantes aos que pregunta.

[CORGA: A. Cortizas, *Leonel*, Xerais, 1996]

E ademais, non estaba ela para poderlle contar o que soñara, e para informala do seu último descubrimento no pergameo da lámpada.

_Pois por aí preto están as minas das que estamos a falar informoulle o avó.

[CORGA: M. Rivas, *O lapis do carpinteiro*, Xerais, 1998]

Asinaba a madre Izarne dende o sanatorio penal de Porta Coeli e informábaa do horario de trens do traslado.

Na orde en que lle comunicaron a custodia do preso, que faría na compañía do cabo Herbal, de volta ao seu destino en Galicia, informábaselle de que se trataba dun "destacado elemento desafecto ao réxime", condenado a cadea perpetua.

Despois do incidente da partida, cando lle informaron que era o médico, adoptara unha actitude de camaradería, pero coa superioridade dun guía de excursionismo.

Non é pintor de brocha, informoulle o sarxento Landesa ao axente finalmente encargado da vixilancia

Informareino, por suposto.²²⁷

Non obstante tamén encontramos algún que outro caso da mesma estrutura cuxo complemento de persoa aparece tanto con pronome de acusativo como co de dativo:

[CORGA: A. Conde, *Sempre me matan*, Ir Indo, 1995]

Ó longo dela Graciano Santillana informouno das ocupacións a manter nos peiraos.

Non acudira alí con aquela intención, pero nese xusto instante sentiuse na obriga de informalo do sucedido entre a súa filla e mailo seu sobriño.

Ó longo dela informouno Celia de que levara con ela as chaves do moble que agachaba a caixa blindada e que, salvo que forzara a pechadura, cousa que non lle aconsellaba facer, posto que o tal moble era unha antigüidade do XVII que lle custara un fortunón ós seus pais, a mellor idea que podería ter sería a de pedir unha copia notarial canto antes.

A ela amosoulle as xoias, tal que se foran trofeos de caza e non dubidou en, no máis grande dos segredos, informarlle do papel que Blas xogara naquela aventura insólita e gozosa.

²²⁷ Incluímolo aquí mais é un caso claramente ambiguo, pois o contexto non nos permite asegurar que a forma de acusativo proceda da pronominalización dun complemento directo + animado (estrutura "informar (a) alguén de algo") ou dun complemento directo - animado (polo tanto, estrutura "informar algo a alguén").

[CORGA: E. Alonso, *Teatro imprevisto*, Tris Tram, 2006]

Federico: Xa está ben, Rosario!, eu civilizadamente quería informarte...

Federico: Ben!, eu só quería informarte, sincerarme contigo, máis nada.

Mónica: Veño dicirche que se queres saber como, cando e porqué... estou disposta a informarte polo miúdo... seguro que as faladurías...

Acoutación: Son apartes como os do teatro clásico, é dicir a personaxe fala directamente co público para informarlle de algo, agora ben, diferéncianse dos do teatro clásico en que aqueles son pequenas frases curtas e rápidas e aquí serán verdadeiros monólogos de duración variable.

2) INSTAR (A) ALGUÉN A / PARA ALGO

a) no corpus de adestramento xornalístico

Pronominalización en forma acusativa:

Non se documenta ningún caso

Pronominalización en forma dativa:

A cantidade obxectada deberase ingresar nalgún proxecto alternativo (non se nega a capacidade recadatoria do Estado, só a recadación para fins militares), e o comprobante deste ingreso inclúese na documentación da declaración do IRPF xunto cunha carta dirixida ó Delegado da Facenda expresándolle os termos da obxección fiscal, e instándolle a que no futuro os impresos inclúan a liña da obxección fiscal para recoñecer e facilitar este dereito. [CORGAetq: *TN1997-06/7*]

Os lexisladores do Consello de Europa acusaron onte a Estados Unidos de "ignorar" os dereitos humanos na súa "guerra contra o terror" e instáronlle a poñer fin a esta situación "inaceptable", incluídas as detencións "ilegais" en Guantánamo. [CORGAetq: *GH2007-03-17/72*]

Sen pronominalizar, complemento sen preposición:

O presidente do Banco Central Europeo (BCE), Wim Duisenberg, instou onte, xoves, os gobernos dos países do euro para que non se relaxen nos seus esforzos para reduci-lo déficit público. [CORGAetq: *CG2000-11-24/50*]

Iraq reaccionou á forte caída dos prezos do cru instando os seus dez socios na OPEP para recortar inmediatamente a oferta por un total dun millón de barrís diarios. [CORGAetq: *CG2001-10-23/40*]

Neste contexto, acusou o Executivo autonómico de "manipular" os datos e instou os seus máximos responsables para dar a cara ante os galegos e a facer públicas as cifras "reais". [CORGAetq: *GH2008-08-05/37*]

Instou o Goberno para "deixarse de escusas" en asuntos que "van máis alá do financiamento" e a "recortar" en gastos de publicidade e propaganda para que as persoas "desesperadas" que se agochan detrás dos datos de Dependencia poidan cobrar as prestacións ás que teñen dereito. [CORGAetq: *GH2008-08-05/37*]

Sen pronominalizar, complemento con preposición:

A Consellería de Industria e Comercio instou a Erimsa para que presentase unha auditoría elaborada "por unha empresa recoñecida polo Ministerio de Industria e colaboradora da Administración en asuntos mineiros" para coñece-la situación da explotación dos xacementos de cuarzo da citada empresa en Galicia. [CORGAetq: *CG1996-02-08/41*]

Así mesmo, instou a Galicia para "ser madura" no plano económico e para "acudir ós mercados de capitais". [CORGAetq: CG2000-12-29/17]

Nunha resolución aprobada na Eurocámara, instase tamén ós bancos da zona euro para que abastezan con billetes de euro de baixa denominación os seus caixeiros automáticos durante os primeiros días do mes de xaneiro". [CORGAetq: CG2001-10-26/33]

b) no corpus de adestramento de ficción

Nos textos seleccionados aleatoriamente para extraer as 200 000 palabras que constitúen o corpus de adestramento de ficción non se rexistra o verbo *instar*, seguido ou non de pronome.

c) nas revistas do CORGA

Pronominalización en forma acusativa:

Non se rexistra ningún caso

Pronominalización en forma dativa:

A cantidade obxectada deberase ingresar nalgún proxecto alternativo (non se nega a capacidade recadatoria do Estado, só a recadación para fins militares), e o comprobante deste ingreso inclúese na documentación da declaración do IRPF xunto cunha carta dirixida ó Delegado da Facenda expresándolle os termos da obxección fiscal, e instándolle a que no futuro os impresos inclúan a liña da obxección fiscal para recoñecer e facilitar este dereito. [CORGA: TN1997-06/7]

Sen pronominalización, complemento sen preposición:

Non se rexistra ningún caso

Sen pronominalización, complemento con preposición:

En 1960 o Ministerio de Vivenda insta ao concello para que presente un Plano Xeral de Urbanismo no prazo de un ano. [CORGA: GI1996-03/31]

A isto cómpre sumar a proposición non de lei presentada polo BNG no Parlamento de Galicia, e aprobada por unanimidade o 24 de maio de 1994, pola cal se instaba á Xunta a reclamar do Goberno do Estado a revocación da concesión feita polo Ministerio de Obras Públicas, Transporte e Medio Ambiente a favor das empresas Electra del Viesgo e Hidroeléctrica del Cantábrico para o aproveitamento do río Navia. [CORGA: GI1996-04/75]

Para superar a polémica diplomática, a Declaración de Salamanca estipulou un rexeite dos xefes de Estado e de goberno "ao bloqueo económico, comercial e financeiro" que Washington

mantén con Cuba, evitando insistir no carácter político e militar da figura do "bloqueo" e instando ao goberno estadounidense a poñer fin a esa práctica. [CORGA: TN2005-11/17]

Xa en Galicia, no mes de febreiro aprobábase no Parlamento unha proposición non de lei do PSdeG para instar ao Goberno da Xunta á creación dunha unidade, tamén no ámbito autonómico, sobre Muller e Ciencia. [CORGA: TN2006-04/6]

Callón considera que o caso galego é semellante ó das comunidades mencionadas, só que en Galiza a situación en vez de producirse con comunidades lindeiras, reproducése entre estados estremeiros, polo que tamén insta ó Goberno galego á negociación desta medida que non só atende a unha semellanza lingüística, senón tamén cultural. [CORGA: CC2007-10/7]

En Galicia, a raíz dunha proposición non de lei presentada polo deputado do BNG Iago Tabarés, o Parlamento instou á Xunta á elaboración dun mecanismo propio, que permitiría entrar en ámbitos de privación da liberdade baixo competencia autonómica, tal é o caso de psiquiátricos ou de centros de menores. [CORGA: TN2008-02/13]

d) nos xornais do CORGA

Pronominalización en forma acusativa:

A Asociación Latinoamericana de Dereitos Humanos (ALDHU) retiroulle o premio 'Monseñor Leónidas Proaño' ó ex secretario xeral da ONU Javier Pérez de Cuellar, pola súa petición ó Goberno e ós militares peruanos de "maior celeridade nas operacións de limpeza de soldados ecuatorianos na fronteira conflictiva", instándoos a "un emprego militar máis efectivo e letal". [CORGA: CG1995-02-07/152]

O Concello de Oleiros publicou un anuncio no que anima os cidadáns a "ve-los grandes documentais da Segunda de TVE" e instaos para que "non perdan o tempo coa tele-lixo". [CORGA: CG1995-02-14/168]

Barbero remitiulle onte mesmo á Sala Segunda do Supremo unha resolución pola que o insta a pedirlle ó Congreso o suplicatorio de Alfonso Guerra, necesario para actuar xudicialmente contra os membros do Parlamento. [CORGA: CG1995-02-23/68]

Esta organización envioulle unha carta ó conselleiro de Pesca, Xoán Caamaño, instándoo a adoptar medidas para atallar este costume. [CORGA: CG1999-05-18/45]

Ademais, instano para comprende-lo proceso posto en marcha para que o pobo basco "decida democraticamente e sen presións de ninguén cómo quere que sexan as relacións entre nosoutros os bascos, así como co resto dos pobos e nacións do mundo". [CORGA: CG1999-05-18/45]

O portavoz de Interior do PP no Congreso, Ignacio Gil Lázaro, pediulle onte ó secretario xeral do PSOE, José Luis Rodríguez Zapatero, "coherencia, prudencia e miras de Estado" ó aborda-

la política antiterrorista e instouno para non "forzar unilateralmente un aparente clima de tensión". [CORGA: CG2000-11-20/30]

A Mesa pola Normalización Lingüística critica que o presidente da Xunta, Manuel Fraga, eludise a súa petición de intervir ante o Concello da Coruña para instalo para cumprila sentenza do Tribunal Supremo sobre o uso do topónimo oficial. [CORGA: CG2000-11-25/36]

Este colectivo enviara unha carta solicitando que Fraga se dirixise formalmente ó alcalde coruñés, Francisco Vázquez, para instalo para cumprir coa sentenza do Supremo e coa lexislación lingüística vixente. [CORGA: CG2000-11-25/36]

Llamazares reclamoulle ó PP tolerancia e flexibilidade e instouno para que abandone a súa "teimosía e intransixencia, porque na súa opinión, a súa responsabilidade é abri-lo acordo ós demais partidos. [CORGA: CG2000-12-19/62]

A Garda Civil envioulles unha circular a tódolos Corpos e Forzas de Seguridade de Lleida na que os insta para extrema-las precaucións ante a posible presenza de etarras na provincia despois de que un dos membros máis buscados da banda terrorista, Gorka Palacios, fose visto na Vall D'Aran hai uns días segundo publicou o diario Segre. [CORGA: CG2000-12-29/52]

Así mesmo, destacou que "existe unha significativa e continua ameaza por parte dos grupos paramilitares que non aceptan o proceso de paz" e instounos para que se desarmen o antes posible. [CORGA: CG2001-10-25/26]

Entón, o presidente da sala dirixiuse ao letrado para informarlle de que a defensa recoñecera a comprensión do galego, polo que o instou a continuar expresándose no noso idioma. [CORGA: ANT2002-11-07/2]

Tamén se cre que o xefe do Goberno español, José María Aznar, presidente de quenda da Unión Europea (UE), quen mantivo antonte unha prolongada conversación telefónica con Sharon, o instou para cancela-la citada operación militar. [CORGA: CG2002-05-13/27]

O concelleiro de Vía Pública do Concello de Xirona, Josep Brugada, pediulle desculpas públicas ó rico empresario metalúrxico E. P. porque o xoves pasado a policía local o instou a abandonar un céntrico lugar ó confundilo cun esmoleiro. [CORGA: CG2002-05-24/7]

Pola súa parte, o ministro Portavoz, Pío Cabanillas, respondeu que estas palabras serven para "clarifica-lo personaxe que é Arzalluz dunha vez por todas", acusouno de ser "avogado defensor de Batasuna" e instouno para que se manifeste contra a banda terrorista nos mesmos termos que o fai contra o Goberno. [CORGA: CG2002-05-25/3]

O ministro Portavoz, Pío Cabanillas, respondeu a estas declaracións considerando que estas palabras serven para "clarifica-lo personaxe dunha vez por todas", acusou ó presidente do PNV de ser "avogado defensor de Batasuna" e instouno para que se manifeste contra a banda terrorista nos mesmos termos que o fan contra o Goberno. [CORGA: CG2002-05-25/3]

Mentres o PP rexeitaba en Madrid a comparecencia de Aznar no Congreso, Fernández Rosende trataba de convencer á oposición de que o Goberno non nega información e instábaos a presentar máis iniciativas no Congreso. [CORGA: *ANT2003-01-23/7*]

Pedrosa asegurou que a actitude do grupo de goberno "non conduce ó diálogo", polo que o instou para que utilice outros métodos, como a mediación, para solucionar os conflitos laborais. [CORGA: *GH2005-02-09/7*]

Nunha rolda de prensa, Rego asegurou que "os galegos están fartos de Feijóo e as súas ocorrencias" e instouno para que explique por que desexa agora a transferencia "e non a quería cando era pública". [CORGA: *GH2007-03-06/14*]

Tanto ó Bloque como ó PSdeG instounos a asumir que o PP "cheira a vitoria", é "o que hai", apostilou. [CORGA: *GH2007-03-26/9*]

Pronominalización en forma dativa:

Por este motivo, catro deputados do Grupo Parlamentario Socialista (Manuel Vázquez, Ceferino Díaz, José Carlos Baños, Ismael Rego) e o seu voceiro, presentaron unha proposición non de lei instándolle á Cámara galega a esixir do Ministerio de Obras Públicas un compromiso público a este respecto. [CORGA: *CG1995-02-10/19*]

A Mesa Nacional de HB decidiu remiti-las dúas copias do vídeo de ETA que estaban no seu poder, tras recibir o venres, á unha e media da tarde, a notificación do xuíz Baltasar Garzón na que se lles instaba a entrega-los exemplares do citado documento audiovisual que posuían. [CORGA: *CG1996-02-18/55*]

O presidente da ANP esixíralle a Sharon a liberación dos preto de 8.000 palestinos presos en cárceres e centros de detención israelís e instaralle a anunciar unha tregua a fin de conseguir que as faccións armadas palestinas proclamen publicamente un alto ó fogo. [CORGA: *GH2005-02-03/43*]

Os lexisladores do Consello de Europa acusaron onte a Estados Unidos de "ignorar" os dereitos humanos na súa "guerra contra o terror" e instáronlle a poñer fin a esta situación "inaceptable", incluídas as detencións "ilegais" en Guantánamo. [CORGA: *GH2007-03-17/72*]

Sen pronominalización, complemento sen preposición [dos 206 casos, recolleemos só un por ano e xornal]:

Así mesmo, acordaron insta-las autoridades a que traballen para que se cumpran as leis. [CORGA: *CG1995-02-01/141*]

A Unión Sindical Agraria Galega (USAG) instou onte a Consellería de Agricultura á urxente convocatoria da mesa da carne, para analizar a suspensión de pagos presentada por Frigsa, ó non poder facer fronte a 300 acreedores. [CORGA: *VG1995-03-06/4*]

Por outra parte, os comerciantes galegos pediránlle á Xunta que inste os concellos para que regule e controle a venda ambulante en Galicia. [CORGA: *CG1996-02-06/48*]

Piqué insta o Sepla para alcanzar un acordo rápido con Iberia [CORGA: *CG1999-05-01/41*]

A Fiscalía Antidroga instará antes de mañá a Audiencia Nacional para que lle reclame a Grecia a extradición do capo galego Laureano Oubiña. [CORGA: *CG2000-11-01/2*]

A moción insta o Executivo a impulsa-la aprendizaxe das linguas cooficiais en tódolos ámbitos administrativos. [CORGA: *CG2001-10-01/11*]

Queren que o Valedor inste o presidente á elaboración dun novo regulamento orgánico e á modificación dos Estatutos do INLUDES, para acadar representación no Consello Rector. [CORGA: *ANT2002-09-26/20*]

Así mesmo, o presidente do Goberno instou os xibaltareños para que comprendan que o mantemento do estatus actual só pode prexudicalos. [CORGA: *CG2002-05-01/12*]

Louzán admitiu que unha saída a esta situación non é doado e instou os responsables políticos do Concello de Pontevedra para que digan que ilegalidades está a cometer Ence. [CORGA: *GH2005-02-07/6*]

O goberno local da Coruña acusou, na sesión plenaria de onte, o portavoz municipal do BNG, Henrique Tello, de instar a Consellería de Industria (xestionada pola formación nacionalista) a que paralizase o polémico elevador panorámico do monte de San Pedro. [CORGA: *DLV2007-05-03/26*]

Por iso, o sindicato instará a Consellería de Educación e os consellos escolares a que fomenten o uso dos nomes de mulleres representativas para denominar os centros educativos de nova creación. [CORGA: *GH2007-03-07/20*]

Neste contexto, acusou o Executivo autonómico de "manipular" os datos e instou os seus máximos responsables para dar a cara ante os galegos e a facer públicas as cifras "reais". [CORGA: *GH2008-08-05/37*]

Sen pronominalización, complemento con preposición [dos 104 casos, recollemos só un por ano e xornal]:

Respecto da falta de eficacia e da necesidade de adoptar medidas que a solucionen, o vicepresidente instou a Ventura Pérez Mariño a que non fixera críticas xeneralizadas e presentase algunha proposta. [CORGA: *CG1995-02-03/81*]

O Parlamento instará á Xunta para que esta demande á Delegación do Goberno do Estado _que é quen ten as competencias nesta materia_ "un coidadoso control do mercado de explosivos e o seu posible uso polas embarcacións pesqueiras, coa colaboración da Garda Civil do Mar". [CORGA: *VG1995-03-08/7*]

O secretario xeral da Seguridade Social, Adolfo Jiménez, instou a tódolos partidos para que concreten nos seus programas electorais as medidas para desenvolve-lo Pacto de Toledo sobre a reforma das pensións, e considerou que, se non o fan, é "por medo". [CORGA: *CG1996-02-12/88*]

Tamén instou ao Goberno a negociar coas partes un reglamento sobre hidrocarburos. [CORGA: *VG1997-11-14/1*]

Sobre a situación da investigación en España, lamentou que se encontra en desvantaxe con respecto a outros países, polo que instou á Universidade e á Administración a esforzarse para supera-las súas carencias. [CORGA: *CG1999-05-11/40*]

A Mesa pídelle ó presidente da Xunta de Galicia que inste a Francisco Vázquez para cumprir coa sentenza do Supremo [CORGA: *CG2000-11-04/11*]

A emisora sinalou que Powell falou por teléfono onte á madrugada con Arafat, presidente da Autoridade Nacional Palestina (ANP), despois de que a administración estadounidense se mostrara preocupada pola continuación da violencia entre israelís e palestinos, e instara a ámbalas partes a multiplica-los esforzos para respecta-lo alto o fogo. [CORGA: *CG2001-10-03/45*]

O seu deputado Ismael Rego calificou a proposta de "cerimonia de distracción" e insistiu en que o único que queda por facer é executar o acordo que se aprobou no mes de Febreiro coa abstención do BNG, unha iniciativa na que se pide á Xunta que inste ao Goberno central sobre a vixencia da proposta de Fraga do ano 1994. [CORGA: *ANT2002-09-26/22*]

O conselleiro lembrou que xa instou "en dúas ocasións" á Confederación Hidrográfica do Sur a que execute "con urxencia" o plan de mellora integral do saneamento de Melilla contemplado no Plan Hidrolóxico Nacional cun investimento de 2.600 millóns de pesetas, "precisamente para evitar[texto omitido] [CORGA: *CG2002-05-07/15*]

A dirección estatal vén instando a Fraga para que nomee un vicepresidente de contado e deixe marcada a liña a seguir no PPdG. [CORGA: *ANT2003-01-16/1*]

Mazuz instou ó Goberno de Israel a suspender a aplicación desa lei que suporía a confiscación dos bens inmobles en Xerusalén Leste de miles de palestinos que viven en Cisjordania. [CORGA: *GH2005-02-02/45*]

Por outra banda, o Congreso aprobou, a pesar da oposición do PSOE, unha proposición non de lei instando ao Goberno a cambiar este ano o modo de elección do presidente e o vicepresidente da CNMV para que estes cargos sexan votados pola Cámara. [CORGA: *DLV2007-05-09/103*]

O alcalde de Ferrol, Juan Juncal, comprometeuse a instar á Xunta de Galicia a cumprir os acordos asinados no Plan de emprego desenvolvido pola Consellería de Pesca. [CORGA:

GH2007-03-02/21]

O senador do PNV Iñaki Anasagasti instou ao etarra a pedir perdón e a facer un "acto de contrición" ante a sociedade, en lugar de dedicarse a exaltar o que o levou ao cárcere.
[CORGA: *GH2008-08-06/68]*

e) na ficción do CORGA

Pronominalización en forma acusativa:

[CORGA: X. Bernárdez Vilar, *No ano do cometa*, Xerais, 1987]

Mais o que non sabía era que acababa de instalo a realizar unha das cousas que máis cobizaba.

[CORGA: L. Vázquez Xil, *A da alba sería*, Edicións do Cumio, 1996]

Quedou Don Quixote suspenso un intre ata que Sancho meneándolle a perna coa man, instouno:

Sorriu pracidamente, como se estivese no secreto, o frade e saíu, "con permiso", para volver súpeto a acompañar ós manchegos ata a presenza de Don Mauro, Santo Padre Abade do mosteiro do Cebadeiro, quen abrazou a Don Quixote contra o abondoso bandullo, deu a bicar a Sancho o anel abacial e instounos a sentar cabo del o pé do lume que gorentaba a ampla cela e xunto a unhas credencias nas que pousaban pratos con peperetadas e garrafas varias.

[CORGA: R. Rei Núñez, *Expediente Artieda*, Xerais, 2000]

_A ti vouche tomar declaración eu _dille, ó tempo que o insta a baixar polos chanzos que conducen ó soto.

Pronominalización en forma dativa:

Non se documenta ningunha ocorrencia.

Sen pronominalizar, complemento sen preposición:

Non se documenta ningunha ocorrencia.

Sen pronominalizar, complemento con preposición:

[CORGA: M. Gándara, *Magog*, Xerais, 1997]

As voces medran, numerosos soldados corren ó longo da ribeira, instando a Xelmírez a desembarcar.

Aproveitando a viaxe do rei a León, a raíña instou ó bispo Bernardo para que se apoderase da Mesquita en nome de Cristo e borrarase as inmundicias do Profeta.

[CORGA: V. Araguas, *A canción do verán*, Ir Indo, 2001]

O certo é que, decatándose da nosa presenza, Cerdeiriña e Arenosa interromperon as súas gargalladas para, petando no ombreiro da súa compañeira, instar a esta ao silencio.

[CORGA: L. Vázquez Xil, *A da alba sería*, Edicións do Cumio, 1996]

Riron con esta saída de Sancho e a teor desto seguiu unha saborosa plática, ata que don Fernán Alonso instou á parella a descansar, pois á noite era gustoso de que ceasen con el e cunha representación de colexiais, convite que a parella aceptou de bo grado.

Referencias dos exemplos de prensa

ANT: A Nosa Terra

C: *Consumer Eroski*

CC: *Código Cero*

CG: O Correo Galego

DLV: De luns a venres

DOG: *Diario Oficial de Galicia*

DX: *Díxitos. Novas do Centro de Supercomputación de Galicia*

EC: *Entregas de Comunicación Cultural*

F: Feiraco

GH: Galicia Hoxe

GI: Galicia Internacional

MC: *Man Común. Revista galega mensual de información xeral*

P: *Petroglifo. Revista da Asociación de Empresarios de Artes Gráficas de Galicia*

TDN: *Teatro do Noroeste*

TE: *Teima. Revista Galega de Información Xeral*

TN: Tempos Novos

VG: La Voz de Galicia

Bibliografía

- Abraham, W. (1981): *Diccionario de terminología lingüística actual*, Madrid, Gredos.
- Aguirre Moreno, J. L.; N. Andiön Rodríguez e X. Gómez Guinovart (2001): “Aspectos ortográficos, léxicos y morfosintácticos del etiquetado lingüístico de un corpus de informática en lengua gallega” en *Procesamiento del Lenguaje Natural*, 27, pp. 13-19.
- Aguirre Moreno, J. L.; A. Álvarez Lugrís e X. Gómez Guinovart (2002a): “Etiquetario morfosintáctico del SLI para corpus de lengua gallega: aplicación al corpus paralelo TECTRA” en *Procesamiento del Lenguaje Natural*, 28, pp. 23-34.
- Aguirre Moreno, J. L.; A. Álvarez Lugrís e X. Gómez Guinovart (2002b): “Aplicación do etiquetario morfosintáctico do SLI ó corpus de traducións TECTRA” en *Viceversa*, 7/8, 2001/2002, pp. 207-229.
- Alarcos Llorach, E. (1994): *Gramática de la lengua española*, Madrid, Espasa Calpe.
- Álvarez, R., X. L. Regueira e H. Monteagudo (1986): *Gramática Galega*, Vigo, Galaxia, 1989, 2ª ed.
- Álvarez, R. e X. Xove (2002): *Gramática da lingua galega*, Vigo, Galaxia.
- Álvarez, R. (1994): “Secuencias de pronomes átonos en galego moderno” en Ramón Lorenzo (ed.), *Actas do XIX Congreso de lingüística e filoloxía románicas*, vol. VI, A Coruña, Fundación Pedro Barrié de la Maza, pp. 247-265.
- Álvarez, R. (1999): “Entre min e ti” en *Homenaxe ó Profesor Camilo Flores*, Volume I, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 117-141.
- Álvarez de la Granja, M. (2003): *As locucións verbais galegas*, Verba, anexo 52, Santiago de Compostela, Universidade de Santiago de Compostela.
- Babarczy, A.; J. Carrol e G. Sampson (2006): “Definitional, personal, and mechanical constraints on part of speech annotation performance” en *Natural Language Engineering*, 12, pp. 77-90. [<http://www.grsampson.net/ADp.html>]

- Barcala, F. M.; M. A. Molinero e E. Domínguez (2005): “Construcción de sistemas de recuperación de información sobre corpóra textuales estructurados de grandes dimensiones”, en *Procesamiento del Lenguaje Natural*, 34, 2005, pp. 41-48.
- Barcala, F. M.; M. A. Molinero e E. Domínguez (2007): “XML rules for enclitic segmentation”, en R. Moreno-Díaz, F. Pichler e A. Quesada-Arencibia (eds.), *Computer Aided Systems Theory*, volume 4739, *Lecture Notes in Computer Science*, Berlin-Heidelberg-New York, Springer-Verlag, 2007, pp. 273-281.
- Bello, A. (1847): *Gramática de la lengua castellana destinada al uso de los americanos*, Santiago de Chile, Imprenta del Progreso. Estudio e ed. de R. Trujillo: *Gramática de la lengua castellana destinada al uso de los americanos. Con las notas de Rufino José Cuervo*, Madrid, Arco/Libros, 1988.
- Borba, F. S. (2002): *Dicionário de usos do Português do Brasil*, São Paulo, Editora Ática.
- Bosque, I. (1991): *Las categorías gramaticales. Relaciones y diferencias*, Madrid, Síntesis.
- Bray, T.; P. Jean e Sperberg-McQueen (eds.): *Extensible Markup Language (XML) 1.0 (Fifth Version)*. [<http://www.w3.org/TR/REC-xml>]
- Buarque de Holanda Ferreira, A. (1999): *Novo Aurélio Século XXI: o dicionário da língua portuguesa*, Río de Xaneiro, Nova Fronteira.
- Burnage, G. e D. Dunlop (1993): “Encoding the British National Corpus”, en J. Aarts, P. de Haan e N. Oostdijk (eds.), *English Language Corpora: Design, Analysis and Exploitation*, Amsterdam, Rodopi.
- Burnard, L. (1997): “The Text Encoding Initiative's. Recommendations for the Encoding of Language Corpora: Theory and Practice” [Seminario de Industrias de la Lengua, Soria, 14-18 de xullo de 1997].
- Burnard, L. e S. Bauman (eds.) (2009): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [<http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>]

- Carballeira Anllo, X. M. (coord.) (2000 e 2009): *Gran Dicionario Xerais da Lingua*, Edicións Xerais de Galicia, Vigo. Edición revisada e aumentada: *Gran Dicionario Xerais da Lingua*, Vigo, Xerais, 2009, 2 v.
- Carballo Calero, R. (1966): *Gramática elemental del gallego común*, Vigo, Galaxia, 5ª ed., 1976.
- Carreras, X.; I. Chao; L. Padró e M. Padró (2004): “FreeLing: An Open-Source Suite of Language Analyzers” en *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisboa. [<http://www.lsi.upc.edu/~nlp/papers/carreras04.pdf>]
- Chacón Calvar, R. e M. Rodríguez Alonso (1993): *Diccionario crítico de dúbidas e erros da lingua galega*, Sada, Edicións do Castro.
- Cidrás Escáneo, F. A. (1991): “As construcións pronominais en galego. Proposta de clasificación e caracterización” en M. Brea e F. Fernández Rei (coords.), *Homenaxe ó Profesor Constantino García*, I, Departamento de Filoloxía Galega, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 101-125.
- Civit Torruella, M.; Castellón Masalles I. e M. A. Martí Antonín (2001): “Creación, etiquetación y desambiguación de un corpus de referencia del español” en *Procesamiento del Lenguaje Natural*, 27, pp. 21-28.
- Civit Torruella, M. (2003): *Criterios de etiquetación y desambiguación morfosintáctica de corpus en español*. Procesamiento del Lenguaje Natural, Monografía 3. [<http://www.sepln.org/monografiasSEPLN/monografiaCivit.pdf>]
- Civit, M.; A. Ageno; B. Navarro; N. Bufi e M. A. Martí (2003b): “Análisis cualitativo y cuantitativo del acuerdo entre anotadores en el desarrollo de corpus interpretados lingüísticamente”, en *Procesamiento del Lenguaje Natural*, 31, pp. 201-208. [<http://www.sepln.org/revistaSEPLN/revista/31/31-Pag201.pdf>]
- Corpas Pastor, G. (1997): *Manual de fraseología española*, Madrid, Gredos.
- Costa Casas, X. X.; Mª A. González Refoxo, C. C. Morán Fraga e X. C. Rábade Castiñeira (1988): *Nova Gramática para a aprendizaxe da lingua*, A Coruña, Vía Láctea.

- Cunha, C. e L. Cintra (1984): *Nova Gramática do Português Contemporâneo*, Lisboa, Edições João Sá da Costa, 2005, 18ª ed.
- Cures Vázquez, L.; X. M. Mosquera Carregal e M. Negro Romero (2005): “Problemas na identificación e acentuación das interrogativas indirectas: *Dime que acentúo ben / Dime qué acentúo ben*” en R. Álvarez e H. Monteagudo (eds.), *Norma lingüística e variación. Unha perspectiva desde o idioma galego*, Santiago de Compostela, Consello da Cultura Galega / Instituto da Lingua Galega, pp. 165-188 (Colección Base. Serie Lingüística; 3).
- Di Tullio, Ángela L. (2008): “Participios y adjetivos” en Mª J. Rodríguez Espiñeira e J. Pena Seijas (coords.) *Categorización lingüística y límites intercategoriales*, Verba, anexo 61, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 99-125.
- Díaz Regueiro, M. (1992): *Os verbos galegos*, Santiago de Compostela, Consellería de Educación e Ordenación Universitaria. Dirección Xeral de Política Lingüística.
- Domínguez Noya, E. (2008): “O Corpus de Referencia do Galego Actual (*CORGA*): presente e futuro” en E. González Seoane, A. Santamarina e X. Varela Barreiro (eds.), *A lexicografía galega moderna: Recursos e perspectivas*, Santiago de Compostela, Consello da Cultura Galega / Instituto da Lingua Galega, pp. 139-151 (Colección Base. Serie Lingüística; 6).
- Domínguez Noya, E.; Fco M. Barcala Rodríguez e M. A. Molinero (2009): “Avaliación dun etiquetador automático estatístico para o galego actual: Xiada” en *Cadernos de Lingua*, 30/31, A Coruña, Real Academia Galega, pp. 151-193.
- Domínguez Noya, E. e X. M. Mosquera Carregal (2011): “Corrector ortográfico especializado para o proxecto IANUS” en X. M. Mosquera Carregal (ed.) *Lingua e Sanidade. VII Xornadas sobre Lingua e Usos*, A Coruña, Universidade da Coruña, pp. 91-123.
- Domínguez Noya, E. (2012): “Partículas exceptivas: problemas de delimitación e proposta de análise”, *Cadernos de Lingua* 34, A Coruña, Real Academia Galega pp. 5-64.

- Domínguez Portela, S. (2006): “Pois: dende a conxunción ó marcador discursivo” en *Cadernos de Lingua*, 28, A Coruña, Real Academia Galega, pp. 145-173.
- Dubois J.; M. Giacomo; L. Guespin; C. Marcellesi; J.B. Marcellesi e J. P. Mével (1992): *Diccionario de Lingüística*, Madrid, Alianza Editorial.
- Fernández Salgado, B. (dir.) (2004): *Diccionario Galaxia de Usos e Dificultades da Lingua Galega*, Vigo, Galaxia.
- Ferreiro, M. (1999): *Gramática histórica galega. I. Fonética e Morfosintaxe*, Santiago de Compostela, Laiovento.
- Ferreiro M. (2002): “Marcadores do discurso e cognición: o caso de *pero*” en X. A. Fernández Roca e M^a. J. Martínez López (coords.), *Vir bonus docendi peritus: Homenaxe a José Pérez Riesco*, A Coruña, Universidade da Coruña, Facultade de Filoloxía, pp. 65-83.
- Freixeiro Mato, X. R. (2000): *Gramática da lingua galega. II. Morfosintaxe*, A Coruña, A Nosa Terra.
- Freixeiro Mato, X. R. (2001): “Interferencia e niveis de uso dalgunhas construcións con pronome posesivo en galego”, *Revista Galega de Filoloxía*, nº 2, pp. 69-88. [<http://hdl.handle.net/2183/2576>]
- Freixeiro Mato, X. R. (2003): *Gramática da lingua galega. IV. Gramática do texto*, A Coruña, A Nosa Terra.
- Freixeiro Mato, X. R. (2005a): “*Non obstante, sen embargo, nembargantes(s)* e outros adverbios conectores do galego escrito. Perspectiva diacrónica e uso actual”, *Verba*, 32, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 247-274.
- Freixeiro Mato, X. R. (2005b): *Os marcadores discursivos. Conectores contraargumentativos no galego escrito*, *Revista Galega de Filoloxía*, monografía 3, A Coruña, Área de Filoloxías Galega e Portuguesa da Universidade da Coruña.
- Frías Conde, X. e X. A. López Silva (2001): “As construcións con *se* en galego: O problema da súa natureza e clasificación” en X. L. Regueira e A. Veiga (eds.),

- Da gramática ao dicionario. Estudos de lingüística galega*, Verba, anexo 49, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 127-142.
- García, C. e A. Santamarina (eds.) (1995): *Atlas Lingüístico Galego. Morfoloxía non verbal*, Volume II, A Coruña, Fundación Pedro Barrié de la Maza.
- García, C. e M. González González (dirs.) (1997): *Diccionario da Real Academia Galega*, A Coruña, Real Academia Galega.
- García, C. e M. González González (dirs.) (2004): *Diccionario castelán-galego da Real Academia Galega*, A Coruña, Fundación Pedro Barrié de la Maza/Real Academia Galega.
- García-Miguel, J. M^a (1995): *Las relaciones gramaticales entre predicado y participantes*, Lalia Series Maior, 2, Santiago de Compostela, Universidade de Santiago de Compostela.
- Gómez Guinovart, X. e S. López Fernández (2009): “Anotación morfosintáctica do Corpus Técnico do Galego” en *Linguamática*, 1, pp. 61-71.
- González González, M.; C. García Mateo; E. Rodríguez Banga e E. Fernández Rei (2002): *Laverca 1.0*, Vigo, Xerais.
- González González, M. e A. Santamarina Fernández (coords.) (2004): *Vocabulario Ortográfico da Lingua Galega*, A Coruña/Santiago de Compostela, Real Academia Galega/Instituto da Lingua Galega.
- González Rei, B. (2004): *Ortografía da lingua galega*, A Coruña, Galinova Editorial.
- Graña Gil, J. (2000): *Técnicas de Análisis Sintáctico Robusto para la Etiquetación del Lenguaje Natural*, Tese de doutoramento, A Coruña, Departamento de Computación, Universidade da Coruña.
- Hermida Gulías, C. (2001): *Ortografía práctica*, Santiago de Compostela, Sotelo Blanco, 2002, 2ª ed.
- Hermida Gulías, C. (2004): *Gramática práctica. (Morfosintaxe)*, Santiago de Compostela, Sotelo Blanco.

- Hermida Gulías, C. (2010): “Lingua e sexismo en dicionarios de galego” en X. M. Mosquera Carregal (ed.), *Lingua e Xénero. VI Xornadas sobre Lingua e Usos*, A Coruña, Servizo de Normalización, Universidade da Coruña, pp. 89-114.
- Ide, Nancy (coord): “Corpus Encoding Standard. Versión 1.5.”
[<http://www.cs.vassar.edu/CES>]
- Instituto da Lingua Galega/Real Academia Galega (1989): *Vocabulario Ortográfico da Lingua Galega. Versión provisional*, [Edición fotocopiada], Santiago de Compostela.
- Instituto da Lingua Galega/Real Academia Galega (1989): *Diccionario inverso da lingua galega (Versión provisional)*, [Edición fotocopiada], Santiago de Compostela.
- Julián Mariscal, O. (2009): “La excepción en la documentación andaluza de Alfonso X” en Daniella Ibba (coord.), *Interlingüística. La lingüística entre los jóvenes investigadores*, Girona, Documenta Universitaria, pp. 82-96.
[<http://access.trencabits.com/~ajl/interlinguistica.php?a=ficha&id=7>]
- Jurafsky, D. e J.H. Martin (2000): *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall.
- Ledo Cabido B. (dir.) (2004): *Diccionario de galego*, Vigo, Ir Indo.
- Leech, G. e A. Wilson (1994): “Morphosyntactic Annotation”, Draft-Work In Progress, EAGLES Document EAG-CSG/IR.T3.1. en N. Calzolari e J. M. McNaught (eds.), *EAGLES Interim Report EAG-EB-IR-2*.
- Leech, G. e A. Wilson (1996): *Recommendations for the Morphosyntactic Annotation of Corpora*. EAG-TCWG-MAC/R.
[<http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>]
- López Martínez, M^a S. (1993): *O complemento directo con preposición a en galego*, Verba, anexo 36, Santiago de Compostela, Universidade de Santiago de Compostela.
- López Viñas, X.; C. Lourenço Mória e M. Moreda Leirado (2010): *Gramática práctica da lingua galega. Comunicación e expresión*, A Coruña, Baía Edicións.

- Manzano Rovira, C. (2006): *La contraargumentación oracional y discursiva en el castellano medieval*, Trballo de Investigación Titorado, Santiago de Compostela, Departamento de Lingua Española, Universidade de Santiago de Compostela.
- Martín Zorraquino, M^a. A. e J. Portolés Lázaro (1999): “Los marcadores del discurso” en I. Bosque e V. Delmonte (dirs.) *Gramática descriptiva de la lengua española*, vol. III, RAE: *Entre la oración y el discurso*. Madrid, Espasa, pp. 4051-4213.
- Martínez Iraolak, D. (2007): *Supervised Word Sense Disambiguation: Facing Current Challenges*, en Procesamiento del Lenguaje Natural, Monografía 5. [<http://www.sepln.org/monografiasSEPLN/monografiaDavid.pdf>]
- Méndez Pazó, F.; F. Campillo Díaz; E. Rodríguez Banga e E. Fernández Rei (2003a): “Análisis morfosintáctico estadístico en lengua gallega” *Procesamiento del Lenguaje Natural*, 31, pp. 159-166. [<http://www.sepln.org/revistaSEPLN/revista/31/31-Pag159.pdf>]
- Méndez Pazó, F.; F. Campillo Díaz e E. Rodríguez Banga (2003b): “Modelado morfológico estadístico aplicado a la síntesis de voz” presentado en *URSI 2003: XVIII Simposium de la Unión Científica Internacional de Radio*. [http://w3.iec.csic.es/ursi/articulos_modernos/articulos_coruna_2003/actas_pdf/SESION%201/S1.%20Aula%202.4/1324%20-%20MODELADO.pdf]
- Mira Mateus, M. H.; A. M. Brito; I. Duarte e I. Hub Faria e S. Frota, G. Matos, F. Oliveira, M. Vigário e A. Villalva (2003): *Gramática da Língua Portuguesa*, Lisboa, Caminho, 5^a ed., revisada e aumentada.
- Moliner, M. (2007): *Diccionario de uso del español*, Madrid, Gredos.
- Monachini, M. e N. Calzolari (eds.) (1996): *Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora. A Common Proposal and Applications to European Languages*. EAG-CLWG-MORPHSYN/R. [<http://www.ilc.cnr.it/EAGLES96/morphsyn/morphsyn.html>]
- Monachini, M. e N. Calzolari (1999): “Standardization in the Lexicon” en H. van Halteren (ed.) *Syntactic wordclass tagging*, Dordrecht, Kluwer Academic Publishers, pp. 149-174.

- Moreda Leirado, M. (2009): “Marcadores conversacionais no galego actual” en *Anuario del Centro de Estudios Gallegos*, Montevideo, Universidad de la República, Facultad de Humanidades y Ciencias de la Educación, pp. 187-207.
- Morel, J.; S. Torner; J. Vivaldi; Ll. de Yzaguirre e M.T. Cabré (1998): *El Corpus de l'IULA: Etiquetaris*, Barcelona, Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 1997; 2ª ed. revisada e corrixida.
- Mosquera Carregal, X. M. (ed.) (2010): *Lingua e Xénero. VI Xornadas sobre Lingua e Usos*, A Coruña, Servizo de Normalización, Universidade da Coruña.
- Mounin, G. (dir.)(1979): *Diccionario de lingüística*, Barcelona, Editorial Labor.
- Muñoz de la Fuente, Mª F. (1996): “Variedad / unidad en la expresión lingüística: La exclusión en gallego-portugués” en J. M. Carrasco González e A. Viudas Camarasa (eds.), *Actas del Congreso Internacional Luso-Español de Lengua y Cultura en la Frontera*, vol. II, Cáceres, Universidad de Extremadura, pp. 47-56.
- Nica, I. (2006): *El conocimiento lingüístico en la desambiguación semántica automática*, en *Procesamiento del Lenguaje Natural*, Monografía 5. [<http://www.sepln.org/monografiasSEPLN/monografiaIuliaNica.pdf>]
- Nicolás Rodríguez, R. (1993): *Diccionario dos verbos galegos*, Pontevedra, Edicións do Cumio.
- Pavón Lucero, Mª. V. (1999): “Clases de partículas: preposición, conjunción y adverbio” en I. Bosque e V. Delmonte (dirs.) *Gramática descriptiva de la lengua española*, vol. I, Real Academia Española: *Sintaxis básica de las clases de palabras*. Madrid, Espasa, pp. 565-655.
- Portolés, J. (2001): *Marcadores del discurso*, Barcelona, Ariel Practicum.
- Real Academia Española (1973): *Esbozo de una nueva gramática de la lengua española*, Madrid, Espasa-Calpe.
- Real Academia Española (1999): *Ortografía de la lengua española*, Madrid, Espasa-Calpe.
- Real Academia Española (2001): *Diccionario de la lengua española*, Madrid, Espasa Calpe, 2 vol., 22ª ed.

- Real Academia Española/Asociación de Academias de la lengua española (2005): *Diccionario panhispánico de dudas*, Madrid, Santillana Ediciones Generales.
- Real Academia Española/Asociación de Academias de la lengua española (2009): *Nueva gramática de la lengua española*, Madrid, Espasa.
- Real Academia Española/Asociación de Academias de la lengua española (2010): *Ortografía de la lengua española*, Madrid, Espasa.
- Real Academia Galega/Instituto da Lingua Galega (1982): *Normas ortográficas e morfolóxicas do idioma galego*, A Coruña/Santiago de Compostela, Real Academia Galega/Instituto da Lingua Galega; 13ª ed. revisada, 1995; 18ª ed. revisada, 2003.
- Reis, R. e J. J. Dias de Almeida (1998): “Etiquetador morfo-sintáctico para o Português” en M.R. Braga Marquillas e M.A. Mota (eds.), *Actas do XIII Encontro Nacional da Associação Portuguesa de Linguística*, vol. II, Lisboa, APL-Colibri, pp. 209-221.
- Rodríguez Rodríguez, A. e M. Davila Ventura (2003): *Lingua galega. Criterios lingüísticos*, Materiais lingüísticos, 2, Vigo, Universidade de Vigo.
- Rojo, G. (1978): *Cláusulas y oraciones*, Verba, anexo 14, Santiago de Compostela, Universidade de Santiago de Compostela.
- Rojo, G.; E. Domínguez Noya e F. M. Barcala Rodríguez (2009): *Léxico de XIADA: Etiquetador/Lematizador do Galego Actual*. [<http://corpus.cirp.es/xiada/descargas>]
- Ruiz Gurillo, L. (2001): *Las locuciones en español actual*, Madrid, Arco Libros.
- Sampson G. e A. Babarczy (2003): “Limits to annotation precision” en *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora (LINC03)*, Budapest, pp. 61-68. [http://www.cl.cam.ac.uk/~ar283/eacl03/workshops03/W03-w3_eacl03sampson.local.pdf]
- Sampson, G. e A. Babarczy (2008): “Definitional and human constraints on structural annotation of English” en *Natural Language Engineering*, 14, pp. 471-494. [http://journals.cambridge.org/download.php?file=%2FNLE%2FNLE14_04%2FS1351324908004695a.pdf&code=260f851f3b7ff3b600062d1042e9273c]

- Sánchez León, F. (1997): “El etiquetado del Corpus de Referencia del Español Actual (CREA)” [Seminario de Industrias de la Lengua, Soria, 14-18 de xullo de 1997]
- Sánchez León, F. (1994): “A Spanish tagset for the CRATER project”. [http://xxx.lanl.gov/PS_cache/cmp-lg/pdf/9406/9406023v1.pdf]
- Schiller, A. e L. Karttunen (1999): “Lexicons for Tagging” en H. van Halteren (ed.) *Syntactic wordclass tagging*, Dordrecht, Kluwer Academic Publishers, pp. 135-147.
- Santalla del Río, M.P. (1995): “Anotación morfosintáctica”, Madrid, RAE [publicación interna].
- Santalla del Río, M.P. (1995): “Propuesta de anotación morfosintáctica” [Versións 1.0, 2.0 e 3.0], Madrid, RAE [publicación interna].
- Santos Río, L. (2003): *Diccionario de partículas*, Salamanca, Luso-Española de Ediciones.
- Seco, M., O. Andrés e G. Ramos (1999): *Diccionario del español actual*, Madrid, Aguilar.
- Seijo Pereiro, L.; A. Martínez Ínsua; F. Méndez Pazó; F. Campillo Díaz e E. Rodríguez Banga (2004): “A Galician Textual Corpus for Morphosyntactic Tagging with Application to Text-to-Speech Synthesis” en M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa e R. Silva (eds.), *LREC 2004: Proceedings of the 4th International Conference on Language Resources and Evaluation (Lisboa, 26-28 May 2004)*, París, European Language Resources Association (ELRA), pp. 1759-1762. [http://gandalf.aksis.uib.no/non/lrec2004/pdf/1111.pdf]
- Silva Domínguez, C. (2002): *Frases nominais con posesivo en galego. Estructura e valores referenciais*, Verba, anexo 50, Santiago de Compostela, Universidade de Santiago de Compostela.
- Ureña López, L. A. (2001): *Resolución de la Ambigüedad Léxica en Tareas de Clasificación Automática de Documentos*, Procesamiento del Lenguaje Natural, Monografía 1. [http://www.sepln.org/monografiasSEPLN/monografiaUrena.pdf]

- Veiga, A. (1986): “Verbo latino e verbo galego. Notas para unha análise comparativa”, *Verba*, 13, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 75-125.
- Veiga, A. (1991): “Sobre a decadencia do subxuntivo en prótases condicionais en galego e español e a subxuntivización da forma *cantara*” en M. Brea e F. Fernández Rei (eds.), *Homenaxe ó profesor Constantino García*, Santiago de Compostela, Universidade de Santiago de Compostela, I, pp. 497-539.
- Veiga, A. (1993): “Sobre a reorganización das oposicións temporais en subxuntivo e subxuntivo irreal na diacronía do verbo hispánico” en R. Lorenzo (ed.), *Actas do XIX Congreso Internacional de Lingüística e Filoloxía Románicas*, Santiago de Compostela, 1989, A Coruña, Fundación Pedro Barrié de la Maza, V, pp. 435-466.
- Veiga, A. (2005): “Falsos castellanismos en gallego. Anotaciones a la luz de la lingüística diacrónica” en A. Rodríguez Barreiro e A. García Lenza (coords.), *Res Diachronicae Virtual 4: El Contacto de lenguas*, pp. 23-50. [http://www.ajihle.org/resdi/docs/Numero4/23_50_viega_alexandre.pdf]
- Vilela, M. (1999): *Gramática da Língua Portuguesa*, Coímbra, Almedina.
- Vilares, M.; J. Graña; T. Araújo; D. Cabrero e I. Diz (1998): “A Tagger Environment for Galician” en *Workshop on Language Resources for European Minority Languages*, Granada.
- Xove, X. (1982): “Sobre a lingua de Cunqueiro. V. O sistema verbal” en *Homenaxe a Álvaro Cunqueiro*, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 308-338.
- Xove, X. (2010): “A voltas coas contraccións: cun e con un” en *Estudos de Lingüística Galega*, 2, Santiago de Compostela, Universidade de Santiago de Compostela, pp. 247-261.
- Xunta de Galicia (2005): *Plan xeral de normalización da lingua galega*, Santiago de Compostela, Xunta de Galicia.

Referencias web

Acesso a corpos/Disponibilização de corpos (AC/DC)

<http://www.linguateca.pt>

Base de datos sintácticos del español actual

<http://www.bds.usc.es>

Corpus AnCora

<http://clic.ub.edu/corpus/ancora>

Corpus de Extractos de Textos Electrónicos MCT/Público (CETEMPúblico)

<http://www.linguateca.pt/CETEMPUBLICO/>

Corpus de Referencia do Galego Actual (CORGA)

<http://corpus.cirp.es/corga>

Corpus de Referencia do Galego Actual etiquetado (CORGAetq)

<http://corpus.cirp.es/corgaetq>

Corpus del español

<http://www.corpusdelespanol.org>

Corpus do Português

<http://www.corpusdoportugues.org>

Corpus de Referencia del Español Actual (CREA)

<http://www.rae.es/creanet.html>

Corpus de Referência do Português Contemporâneo (CRPC)

<http://alfclul.clul.ul.pt/CQPweb/crpc/index.php>

Corpus Lingüístico da Universidade de Vigo (CLUVI)

<http://sli.uvigo.es/CLUVI/>

Corpus literario TECTRA inglés-galego (TECTRA)

<http://sli.uvigo.es/CLUVI/#tectraig>

Corpus Técnico do Galego (CTG)

<http://sli.uvigo.es/CTG/>

Corpus Técnico Anotado do Galego (CTAG)

<http://sli.uvigo.es/CTAG/>

CRATER Multilingual Aligned Annotated Corpus

<http://www.comp.lancs.ac.uk/linguistics/crater/corpus.html>

Tesouro Informatizado da Lingua Galega (TILG)

<http://ilg.usc.es/tilg/>

Tesouro Informatizado da Lingua Galega (TILG)

<http://sli.uvigo.es/TILG/>

Tesouro Medieval Informatizado da Lingua Galega (TMILG)

<http://ilg.usc.es/tmilg/>

Diccionario da Real Academia Galega (DRAG 2012)

<http://www.realacademiagalega.org/diccionario#inicio.do>

Diccionario de galego (Ir Indo)

<http://www.digalego.com>

Diccionario de partículas discursivas del español

<http://www.dpde.es/>

Etiquetador/Lematizador do Galego Actual (XIADA)

<http://corpus.cirp.es/xiada>

Vocabulario ortográfico da lingua galega (VOLGa)

<http://www.realacademiagalega.org/recursos-volg>