



UNIVERSIDADE DA CORUÑA



Universidade de Vigo

Facultade de Filoloxía USC

Mestrado Interuniversitario en Lingüística Aplicada

Traballo de Fin de Mestrado

**Os problemas léxico-semánticos e a súa anotación no
*Corpus de textos galegos escritos por estudantes no
ámbito académico (CORTEGAL)***

Autora: Natalia Lodeiros Souto

Titora: María Álvarez de la Granja

Xullo de 2020



UNIVERSIDADE DA CORUÑA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA



Universidade de Vigo

Facultade de Filoloxía USC

Mestrado Interuniversitario en Lingüística Aplicada

Traballo de Fin de Mestrado

**Os problemas léxico-semánticos e a súa anotación no
*Corpus de textos galegos escritos por estudantes no
ámbito académico (CORTEGAL)***

Autora: Natalia Lodeiros Souto

Titora: María Álvarez de la Granja

Visto e prace:

Xullo de 2020

ÍNDICE

Resumo	5
1 Introdución.....	5
2 Os corpus de aprendentes con anotación informatizada de erros	6
2.1 Presentación xeral	6
2.2 Análise de Erros e Análise de Erros por Computador	9
2.3 Sistemas de clasificación dos erros	12
2.3.1 Caracterización xeral	12
2.3.2 Sistemas de clasificación de erros nalgúns corpus de aprendentes	13
2.4 Sistemas de clasificación dos problemas léxicos	20
3 O <i>Corpus de textos galegos escritos por estudantes no ámbito académico</i> (CORTEGAL)	23
4 Consecuencias léxicas do contacto do español co galego	26
5 Metodoloxía de traballo	28
5.1 Selección dos textos.....	28
5.2 Selección dos problemas	29
5.3 Clasificación establecida	31
6 Análise dos textos	32
6.1 Problemas léxicos	33
6.1.1 Importacións léxicas do español non adaptadas (L_sub_imp).....	33
6.1.2 Importacións léxicas do español adaptadas (L_sub_impa).....	39

6.1.3	Importacións léxicas do español híbridas (L_sub_imph)	41
6.1.4	Calcos (L_sub_c)	42
6.1.5	Palabras galegas desviadas da norma (L_sub_pgal)	43
6.1.6	Creacións (L_sub_crea)	44
6.1.7	Creacións por analogía (L_sub_creaanal).....	45
6.1.8	Estranxeirismos non necesarios (L_sub_estranx)	46
6.2	Problemas semánticos.....	47
6.2.1	Calcos semánticos (S_sub_csem)	47
6.2.2	Palabras estándar mal empregadas (S_sub_pme)	48
7	Conclusións	50
8	Referencias bibliográficas	55

ÍNDICE DE ILUSTRACIÓN

Ilustración 1.	Tipoloxía dos erros no CUTe. Fonte: Cardoso et al. (2014: 173)	15
Ilustración 2.	Niveis de anotación en MERLIN. Fonte: MERLIN Project (2014: 12)...	19
Ilustración 3.	Erros gramaticais e ortográficos en MERLIN. Fonte: MERLIN project (2014: 28)	20
Ilustración 4.	Erros léxicos no ASK corpus. Fonte: Tenfjord et al. (2006: 1822).....	22
Ilustración 5.	Preguntas das probas ABAU. Fonte: Álvarez de la Granja (2018: 56)....	24

ÍNDICE DE TÁBOAS

Táboa 1.	Importacións léxicas do español non adaptadas	38
Táboa 2.	Importacións léxicas do español adaptadas	40

Táboa 3. Importacións léxicas do español híbridas	41
Táboa 4. Calcos	42
Táboa 5. Palabras galegas desviadas da norma	43
Táboa 6. Creacións	44
Táboa 7. Creacións por analoxía	45
Táboa 8. Estranxeirismos non necesarios	47
Táboa 9. Calcos semánticos.....	47
Táboa 10. Palabras estándar mal empregadas	49

ÍNDICE DE GRÁFICOS

Gráfico 1. Problemas léxicos e semánticos por lemas.....	50
Gráfico 2. Problemas léxicos e semánticos por formas.....	50
Gráfico 3. Problemas léxicos por lemas	51
Gráfico 4. Problemas léxicos por formas	51
Gráfico 5. Importacións léxicas do español por lemas.....	52
Gráfico 6. Importacións léxicas do español por formas	52
Gráfico 7. Problemas semánticos por lemas.....	53
Gráfico 8. Problemas semánticos por formas	53

Resumo

Hoxe en día a Lingüística de Corpus é unha peza fundamental da Lingüística Aplicada. É difícil concibir a lingüística actual sen os corpus. Neste senso, os corpus de aprendentes son ferramentas esenciais para comprendermos e mellorarmos a teoría da adquisición das linguas e as metodoloxías necesarias para o proceso de ensino-aprendizaxe.

Neste traballo afondaremos no funcionamento dos corpus de aprendentes con anotación informatizada de erros e nas súas clasificacións dos problemas léxico-semánticos. Logo, aplicaremos as nosas pescudas sobre o *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL) analizando os problemas léxico-semánticos e ofrecendo unha proposta de clasificación e etiquetaxe.

1 Introducción

Os obxectivos principais deste traballo son elaborar unha proposta de clasificación e anotación dos problemas léxicos e semánticos atopados no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL) e ofrecer unha caracterización dos problemas principais deste tipo. Para isto, primeiro faremos unha presentación xeral dos corpus de aprendentes con anotación informatizada de erros (vid. 2), prestando especial atención á metodoloxía de Análise de Erros por Computador e aos sistemas de clasificación de erros de varios corpus. Unha vez visto o funcionamento destes corpus, presentaremos o corpus CORTEGAL, o seu sistema de anotación e a súa clasificación xeral de problemas (vid. 3), que serán a base da nosa clasificación.

Antes de propoñer a clasificación trataremos as consecuencias do contacto do español co galego no ámbito léxico (vid. 4), imprescindible para comprender o elevado número de unidades non estándares presentes en CORTEGAL e, daquela, para establecer a clasificación dos problemas. Despois, presentaremos a metodoloxía de traballo (vid. 5) coa selección dos textos de análise, os problemas e a súa clasificación.

A continuación, ofreceremos unha análise dos textos (vid. 6) na que presentaremos os problemas atopados clasificados e faremos unha análise cuantitativa e cualitativa dos

resultados. Por último, incluiremos unha breve conclusión (vid. 7), presentando tamén algunha das dificultades encontradas á hora de establecer a clasificación.

2 Os corpus de aprendentes con anotación informatizada de erros

2.1 Presentación xeral

O primeiro no que nos temos que deter é na definición de corpus de aprendentes. De xeito sinxelo e rápido, podemos definir un corpus de aprendentes como unha colección electrónica de textos producidos por aprendentes de lingua. Segundo Granger (2008: 259), tales corpus poden cumprir dúas funcións diferentes, ou ben contribuír a mellorar a teoría da adquisición de segundas linguas sobre a descrición da interlingua e os factores que inflúen nela, ou ben contribuír a desenvolver ferramentas e métodos pedagóxicos para mellorar as necesidades dos aprendentes. Este tipo de corpus aparecen nos anos 90, logo de que xa se viñeran empregando outra clase de corpus na investigación da Adquisición da Linguaxe desde os anos 70. A investigación de Corpus de Aprendentes Informatizados (CAI) nace coa publicación en 1998 de *Learning English on Computer*, editado por Granger (Sánchez Rufat, 2015: 192-195).

Coma todo corpus, os corpus de aprendentes deben estar deseñados seguindo unha serie de criterios estándar. Lozano e Mendikoetxea sintetizan os xa propostos por Sinclair (recollidos en Sánchez Rufat, 2015: 193). Estes criterios son:

1. Débese seleccionar o contido de acordo con criterios externos, como a función comunicativa dos textos e non a linguaxe empregada.
2. O estado de lingua seleccionado para o corpus debe ser representativo.
3. Débese contrastar o corpus cun corpus de control deseñado para esa fin.
4. A estrutura debe ser sinxela.
5. Débense almacenar por separado as etiquetas e o texto en bruto.
6. Os textos deben comprender actos de fala completos, independentemente do tamaño.
7. O deseño e a composición deben estar documentados.
8. Debe existir un equilibrio entre as mostras de lingua oral e escrita se o corpus quere incluír ambas as variedades.
9. O control do tema nun corpus só debe impoñerse polo uso de criterios externos.

10. Os textos deben ser homoxéneos, débense descartar aqueles que non son representativos da variedade seleccionada.

O máis característico destes corpus é o tipo de datos que recollen, xa que estes proveñen de aprendentes. En sentido estrito, o que se recoñece como aprendentes ou *language learners* son falantes que están aprendendo unha lingua que non é a súa L1 nin é unha lingua oficial no seu país ou rexión (Granger, 2008: 260). Esta definición pechada de aprendente de lingua suscita polémica á hora de poder considerar ou non textos escritos na L1 como elementos dun corpus de aprendentes.

Hoxe en día, hai autores como Abel *et al.* (2014) que tamén consideran aprendentes as persoas que están en proceso de adquisición de habilidades dunha L1, por exemplo habilidades tan importantes como a escritura ou o proceso de produción de textos. Nós, consideraremos aprendente de linguas a toda persoa que estea en proceso de aprendizaxe dunha lingua, sexa esta L1 ou L2. Ademais, desde a perspectiva computacional non varía traballar con producións de L1 ou L2, “L1 and L2/FL learner corpora are fully equivalent since both are compilations of textual data that may deviate from the standard variety” (Abel *et al.*, 2014: 2414).

Outra dificultade é o tipo de textos ou datos que se consideran aptos para un corpus de aprendentes. A autenticidade é un trazo definitorio da información de todo tipo de corpus. No caso da produción de textos por parte dos aprendentes a naturalidade varía segundo o propósito. Non é o mesmo facer un exercicio de encher ocos que redactar un texto argumentativo sobre un tema. Para este tipo de corpus é mellor escoller textos que proveñan de *open-ended types of tasks*, é dicir, tarefas nos que os aprendentes teñan que redactar de xeito máis ou menos libre (Granger, 2008: 261).

Granger (2008: 261-263), partindo de corpus de aprendentes de inglés establece unha tipoloxía enfrontada por pares para este tipo de corpus.

- Comercial fronte a académico

Esta diferenza radica en se o corpus está iniciado por unha empresa editorial ou por unha institución educativa. Os máis habituais son os académicos aínda que os comerciais adoitan ser máis grandes.

- Grande fronte a pequeno

Hoxe en día, o formato electrónico permite recoller gran cantidade de información e traballar con corpus de millóns de palabras. A facilidade á hora de recompilar datos axuda a ofrecer un corpus representativo e xeneralizar os resultados. Aínda así, isto non implica que os corpus pequenos non teñan valor. Todo dependerá dos criterios e dos obxectivos de cada corpus.

- Inglés fronte a non-inglés

Coma en todo o universo da tecnoloxía e da investigación lingüística actual o mundo anglófono ten un gran peso no estudo dos corpus de aprendentes. Mais é certo que nos últimos anos os proxectos de corpus de aprendentes noutras linguas diferentes ao inglés están en aumento.

- Lingua escrita fronte a lingua oral

Existen máis corpus de aprendentes que recollen lingua escrita ca lingua oral, pois recompilar textos da lingua oral para un corpus supón, *a priori*, un maior traballo de tratamento dos datos. Normalmente o tipo de lingua escrita que se recolle nestes corpus ten uns fins académicos marcados, pero na actualidade, o desenvolvemento da tecnoloxía da información e da comunicación permite recoller variedades de lingua en contexto de lingua estranxeira ou L2 máis alá do obxectivo académico.

- Diacrónico fronte a sincrónico

Os corpus de aprendentes sincrónicos (*cross-sectional*) recollen información de diferentes tipos de aprendentes nun punto determinado do tempo ou do proceso de aprendizaxe. Pola contra, os corpus diacrónicos ou lonxitudinais, menos frecuentes, recollen información sobre os mesmos aprendentes durante todo o proceso de aprendizaxe. Os investigadores interesados no desenvolvemento da interlingua apostan por corpus quasi-lonxitudinais, é dicir, son corpus sincrónicos que recollen información do estado da lingua de aprendentes de diferentes niveis.

- Para uso pedagóxico directo fronte a indirecto

Nos corpus para o uso pedagóxico os aprendentes son os produtores da información recollida, pero tamén os usuarios e beneficiarios desta. Os profesores recollen esta información nas aulas a partir das actividades.

Pola contra, os corpus que non están deseñados directamente para o uso pedagóxico non se empregan como material de aprendizaxe ou ensino. Estes corpus teñen como obxectivo describir a interlingua dun determinado nivel ou deseñar ferramentas pedagóxicas para aprendentes do mesmo nivel que os produtores da información do corpus.

- Anotados fronte a non anotados

Outra das grandes diferenzas tipolóxicas que nos podemos atopar dentro do mundo dos corpora e, en concreto, nos corpus de aprendentes son os corpus non anotados fronte aos anotados. Que un corpus estea anotado quere dicir que, ademais de presentar os propios textos recollidos, achega varios tipos de información lingüística. Pódese falar da anotación gramatical e da anotación dos erros. A anotación faise a través de ferramentas como lematizadores, analizadores sintácticos e etiquetadores (Granger, 2008: 265). Para falarmos da anotación informatizada de erros aplicada aos corpus de aprendentes, cómpre falar da metodoloxía da Análise de Erros e da súa evolución actual á Análise de Erros por Computador.

2.2 Análise de Erros e Análise de Erros por Computador

Como indica o *Diccionario de términos clave de ELE*, a Análise de Erros (AE) nace nos anos 70 apoiándose nas investigacións de Chomsky sobre a adquisición da lingua materna. Recordemos que este autor consideraba a adquisición coma un proceso creativo no que os nenos constrúen a gramática da lingua en cuestión a partir da exposición á lingua real. O método desta análise seguía os pasos propostos por S. P. Corder (principal representante da AE) en 1997: identificar os erros no seu contexto, clasificalos e describilos, explicar a súa orixe buscando os mecanismos e as fontes de cada un e por último, avaliar a súa gravidade e buscar un posible tratamento.

Dagneaux *et al.* (1998: 164-165) fan un repaso das limitacións deste método e resaltan cinco puntos débiles. A AE está baseada en información heteroxénea de aprendentes, as categorías que establece son difusas, non pode atender a fenómenos como a evitación,

limitase a remarcar o que o aprendiz non pode facer e dá unha imaxe estática da aprendizaxe dunha L2.

Tanto a información heteroxénea como as categorías difusas derivan de erros na metodoloxía (Dagneaux *et al.*, 1998: 164). Hai que procurar reunir mostras ben definidas da lingua dos aprendentes para poder facer logo unha boa clasificación dos erros. As categorías non adoitaban estar ben definidas xa que mesturaban criterios e niveis de análise (descrición e explicación) e ademais eran bastante subxectivas.

En canto á imposibilidade de explicar o fenómeno da evitación e a insistencia en determinadas formas correctas, está determinada polo propio obxectivo do método, “[these] limitations have to do with the scope of EA. EA's exclusive focus on overt errors means that both non-errors, i.e. instances of correct use, and non-use or underuse of words and structures are disregarded” (Dagneaux *et al.*, 1998: 164). É dicir, este método de análise só se centra nas formas que considera erros.

Os avances tecnolóxicos e a propia evolución da lingüística de corpus permitiron a aparición da “anotación de erros” nos textos. Para isto, desenvolvéronse diferentes sistemas e ferramentas. Granger (2008: 266) mostra, por exemplo, o sistema da Cambridge University Press para o inglés, que emprega un código de dúas letras na que a primeira representa o tipo de erro e a segunda identifica a clase de palabra, e o sistema de Louvain para o francés como lingua estranxeira, con tres niveis, dominio do erro (forma, morfoloxía, gramática...), categoría (número, tempo verbal...) e clase de palabra da forma mal empregada.

Normalmente, a anotación de erros tamén inclúe a corrección,

which means that a researcher has the option to start from a particular correction and find the full range of forms that have given rise to that correction or conversely, to retrieve all the possible corrections of a particular erroneous item (Granger, 2008: 266).

Isto quere dicir que, a partir dun erro nun texto do corpus podemos atopar todos os demais erros relacionados con esa forma ou todas as posibles correccións dun erro determinado.

Todo este proceso de detectar, corrixir e anotar o erro encerra un alto grao de subxectividade. Por iso é importante ter unha tipoloxía de erros coherente e sistemática que se poida consultar nun manual, onde todas as etiquetas e as diferentes categorías estean ben definidas e ilustradas (Granger, 2008: 266).

A Análise de Erros por Computador (AEC) ten as súas orixes na AE, pero, como xa indicamos, esta metodoloxía presentaba unha serie de limitacións. Unha delas era a propia información coa que se traballaba, pois esta recollíase como unha lista de erros que non tiña en conta variables como o tipo de tarefas, os aprendentes ou o nivel de lingua, achegando así datos moi heteroxéneos e difíciles de clasificar (Díaz-Negrillo e Fernández Domínguez, 2006: 85), é dicir, existía un problema á hora de sistematizar os datos. Ademais, outra das limitacións que presentaba era a propia clasificación dos erros: este método establecía taxonomías a partir de categorías, bastante subxectivas, de diferentes dominios, superpoñendo criterios.

A AEC supera a limitación da información cos corpus de aprendentes, grazas a que estes achegan unha imaxe clara da lingua do aprendente e permiten recompilar gran cantidade de información e traballar con ela coa axuda de ferramentas informáticas. Ademais, a información que recollen segue uns criterios sistemáticos que axudan a que os datos sexan representativos, ofrecéndoos en contexto real, o que permite comprender mellor o proceso de produción dos aprendentes (Díaz-Negrillo e Fernández Domínguez, 2006: 85). Isto é o que apunta Granger (1998: 268):

Computer-aided error analysis (CEA) involves analyzing learner errors on the basis of learner corpora in which error tags and possible corrections have been inserted with help of a purpose-built editing tool. It differs from previous error analysis studies in some major respects, not least of which is the fact that errors are not isolated from texts in which they originated, as was the case in traditional EA, but rather studied in context alongside cases of correct use and over- and underuse.

En canto á segunda limitación, os investigadores que traballan con este método intentan superar as problemas das clasificacións dos tipo de erro da AE reformulando as taxonomías. Estas taxonomías deberían basearse na descrición dos datos e incluír categorías lingüísticas ben definidas para reducir a subxectividade á hora de clasificar

os erros (Díaz-Negrillo e Fernández Domínguez, 2006: 85-86). Aínda así, a este método tamén se lle achaca a escasa estandarización das taxonomías,

the diversity of error tagging systems seems to be evidence of the constant questioning of emerging approaches to error annotation, and also of the need for a benchmark for the analysis of computerized learner errors (Díaz-Negrillo e Fernández Domínguez, 2006: 86).

2.3 Sistemas de clasificación dos erros

2.3.1 Caracterización xeral

Elaborar un sistema de clasificación de erros consiste en deseñar unha taxonomía que faga referencia ás categorías de erros para que logo os investigadores inclúan estas etiquetas de xeito manual, axudados de editores, nos corpus. Estas ferramentas asocian as etiquetas ás diferentes categorías de erros e permiten ir inseríndoas a medida que se vai revisando o texto. Ademais, moitos destes sistemas tamén permiten introducir correccións. Unha vez que todos os erros están etiquetados, estes poden ser recuperados con softwares de recuperación especializados. En moitos casos o sistema está asociado ao corpus para o que se crea, polo que o propio corpus determina as características da anotación.

Díaz-Negrillo e Fernández Domínguez (2006: 88-91) apuntan e repasan os principios, xa analizados por Granger, para un bo sistema de anotación de erros. Estes sistemas deben ser coherentes, informativos, flexibles e reutilizables.

Para que un sistema de anotación de erros sexa coherente é necesario elaborar unha boa guía de etiquetaxe que inclúa información sobre as categorías de erro e os principios do proceso de anotación. Isto axuda a establecer unhas etiquetas claras que non caian na subxectividade. Aínda así, ao estarmos diante dun proceso manual –as persoas que anotan escollen a etiqueta– segue a existir un certo grao de subxectividade, problema que xa se lle achacaba ao método da Análise de Erros.

Que un sistema sexa informativo indica que as categorías de erro están ben descritas. Canto máis detalladas sexan estas definicións máis exacta será a análise e menos pé deixaremos á subxectividade. Para avaliar o nivel de detalle das etiquetas tense en conta o número de etiquetas que contén o sistema e o tipo de dimensións que cobre, así como

os aspectos que sinala cada etiqueta. Normalmente, os niveis lingüísticos cubertos nestas taxonomías son a ortografía, a gramática e o léxico. Os erros de fonética, pragmática e discurso non sempre están presentes.

A flexibilidade implica que as categorías poden adaptarse segundo os obxectivos das investigacións. Para Granger (apud Díaz-Negrillo e Fernández Domínguez, 2006), a flexibilidade debería ser inherente ás dúas etapas de uso das etiquetas: a anotación e a recuperación. Na anotación deben ser posibles diferentes niveis de especificidade. Do mesmo modo, na etapa de recuperación o sistema debería permitir buscas nos diferentes niveis de especificidade anotados.

Para que o sistema poida ser reutilizable, o conxunto de etiquetas debería permitir o seu uso noutros corpus de aprendentes de diferentes linguas. Por isto, son necesarias categorías de erro xerais. Este principio pretende garantir a reutilización, pero tamén a normalización e estandarización da anotación de erros. Con todo, a maioría dos corpus anotados presentan un sistema con etiquetas diferentes,

The tagset builders may want to make their tagset specific for a given corpus, which is valid and very much depends on the researcher's interests, but they also seem to be aware of the restrictions of reusability of their tagset (Díaz Negrillo e Fernández Domínguez, 2006: 91).

Por último, outro principio destes sistemas de anotación de erros, que recollen Díaz-Negrillo e Fernández Domínguez (2006: 91), é o xa discutido por varios autores, a necesidade de incluír descrições alternativas dos erros ou problemas para prestar atención ás diferentes interpretacións que varias persoas poidan facer da mesma forma e así evitar un alto grao de subxectividade.

2.3.2 Sistemas de clasificación de erros nalgúns corpus de aprendentes

Nas seguintes páxinas presentaremos varios corpus de aprendentes e mostraremos os seus sistemas de clasificación de erros. Dedicaremos un apartado específico para tratarmos a clasificación dos problemas léxicos (vid. 2.4).

KoKo é un corpus de aprendentes de alemán como L1. Este corpus foi creado para investigar e describir as habilidades de redacción do alumnado na fin da secundaria a

través da análise de textos auténticos producidos na aula. A construción deste corpus estivo guiada por dous obxectivos principais: describir as habilidades de redacción dos aprendentes que van acceder á universidade e determinar factores externos que inflúen na distribución destas habilidades, como a rexión, a idade, o xénero etc. Recolleron textos argumentativos sobre o mesmo tema e cuestionarios sobre aspectos sociolingüísticos de 1511 aprendentes (Abel *et al.*, 2014: 2415-2416).

Este corpus presenta dúas dimensións na anotación de erros, por un lado estaría a anotación da categoría lingüística do erro (ortografía, gramática, léxico ou aspectos do nivel textual) e por outro lado a clasificación da modificación con respecto ao estándar (omisión, adición...). Ademais, tamén introducen a corrección do erro. Por exemplo, a clasificación de erros ortográficos contén 28 categorías que se poden clasificar en 7 categorías superiores (Abel *et al.*, 2014: 2417).

O esquema da anotación de erros gramaticais é o seguinte (Abel *et al.*, 2014: 2418):

- 1) Relacións de correspondencia
 - a) Correspondencia: selección errónea dun caso, número, xénero ou persoa dunha palabra con respecto á concordancia
 - b) Referente da correspondencia: a palabra ou frase á que fai referencia o erro de concordancia
- 2) Flexión: formas de flexión incorrectas que son independentes dun elemento gobernante
- 3) Incompleto: oración e frase incompletas, así como o uso incorrecto de elipses
- 4) Redundancia: repeticións erróneas de palabras e partes de frases
- 5) Anacoluto: mestura agramatical de frases e cláusulas
- 6) Orde de palabras: violacións de calquera tipo de restrición da orde de palabras
- 7) Erro gramatical non categorizable

Ademais, cada unha destas anotacións gramaticais presentan esquemas con subclasificacións.

O CUTE, *Corpus of Portuguese Undergraduates' Texts*, é “o primeiro corpus de aprendizagem de escrita académica em português” (Cardoso *et al.*, 2014: 170). Este corpus inclúe 480 producións textuais de estudantes da licenciatura da ESELx (Escola

Superior de Educação de Lisboa). Os textos pertencen a dous xéneros académicos: o artigo de divulgación e o artigo de opinión.

En canto á anotación de erros, este corpus presenta unha tipoloxía específica que ten como base a análise lingüística dos textos considerando os erros dos niveis micro e macroestruturais. Estes están organizados en dous niveis de descrición e anotación: nivel de análise e categoría (Cardoso *et al.*, 2014: 173).

Nível de análise	Categoria
Ortografia	Acentuação
	Maiúsculas/minúsculas
	Translineação
	Hifenização
	Processamento e representação gráfica de fonemas
	Acordo ortográfico
	Outros
Pontuação	Uso da vírgula
	Uso de outros sinais de pontuação
Morfologia	Flexão
	Derivação
Sintaxe	Ordem de palavras
	Topicalização
	Coordenação
Morfosintaxe	Concordância
	Segmentação de palavras
Sintaxe/semântica	Seleção categorial
	Seleção semântica
	Seleção de modo
	Outras incompatibilidades categoriais
	Outras incompatibilidades semânticas
Semântica	Referência nominal
	Referência temporal-aspetual
	Conectores
Gralhas	

Ilustración 1. Tipoloxía dos erros no CUTe. Fonte: Cardoso et al. (2014: 173)

Como se pode observar na táboa, esta tipoloxía contén oito niveis de análise que recollen erros de convencións de representación gráfica (ortografía, puntuación), erros dos diferentes niveis de análise lingüística (morfoloxía, sintaxe, semântica), erros en categorías mixtas (morfosintaxe, sintaxe/semântica) e unha categoría de grallas (Cardoso *et al.*, 2014: 174).

Como vemos na táboa, o nivel de ortografía recolle erros de acentuación, maiúsculas/minúsculas, separación silábica ao final de liña, procesamento e representación gráfica de fonemas, acordo ortográfico e outros; o nivel da puntuación sinala erros de uso da coma e doutros usos dos signos de puntuación; o nivel da morfoloxía, erros de flexión e derivación; o nivel da sintaxe, orde de palabras, topicalización e coordinación; o nivel da morfosintaxe, erros de concordancia e de

segmentación de palabras; o nivel da sintaxe/semántica, erros de selección da categoría, erros de selección semántica, erros de selección de modo verbal e outras incompatibilidades de categoría e semánticas e o nivel da semántica, erros de referencia nominal, de referencia temporal-aspectual e uso de conectores.

O *Cambridge Learner Corpus* é un corpus de aprendentes de inglés como L2 que presenta textos recollidos das probas de exame de Cambridge. Segundo Nicholls (2003: 572), o obxectivo deste corpus é anotar todos os erros dun tipo para que logo se poidan extraer e ser empregados con diferentes fins (docencia, investigación, elaboración de dicionarios...). O sistema de anotación emprega uns códigos que marcan os contextos onde se producen os erros. Estas etiquetas insírense de xeito manual nos textos:

<#CODE>wrong word|corrected word</#CODE>

A maioría dos códigos son de dúas letras, a primeira indica o tipo xeral de erro e a segunda o tipo de palabra ao que se refire. Así, este sistema de clasificación recolle (Nicholls, 2003: 573-575):

- Erros xerais
 - o F wrong Form used (formas mal empregadas)
 - o M something Missing (falta algo)
 - o R word or phrase needs Replacing (a palabra ou a frase necesita ser substituída)
 - o U word or phrase is Unnecessary (a palabra ou frase é redundante)
 - o D word is wrongly Derived (a palabra está mal derivada)
- Tipos de palabras
 - o A Pronoun
 - o C Conjunction
 - o D Determiner
 - o J Adjective
 - o N Noun
 - o Q Quantifier
 - o T Preposition
 - o V Verb
 - o Y Adverb

- Erros de puntuación (tipo de erro + P)
 - MP punctuation Missing (falta signo de puntuación)
 - RP punctuation needs Replacing (corresponde outro signo)
 - UP Unnecessary punctuation (puntuación innecesaria)
- Erros de substantivos contables/incontables (C + tipo de palabra)
 - CN countability of Noun error
 - CQ wrong Quantifier because of noun countability
 - CD wrong Determiner because of noun countability
- Erros de “falsos amigos” (FF + tipos de palabra)
- Erros de concordancia (AG + tipo de palabra)
 - AGA Anaphoric (pronoun) agreement error
 - AGD Determiner agreement error
 - AGN Noun agreement error
 - AGV Verb agreement error
- Outro tipo de erros
 - AS incorrect Argument Structure
 - CE Compound Error
 - CL CoLlocation error
 - ID IDiom error
 - IN Incorrect formation of Noun plural
 - IV Incorrect Verb inflection
 - L inappropriate register (Label)
 - S Spelling error
 - SA American Spelling
 - SX Spelling confusion error TV wrong Tense of Verb
 - W incorrect Word order
 - X incorrect formation of negative

O COPLE2 é un corpus de aprendentes de portugués como L2 e lingua estranxeira que ten diferentes L1 e niveis. Como indican Río e Mendes (2018: 227), a anotación neste corpus faise coa plataforma TEITOK, a mesma que emprega CORTEGAL. Ten dous sistemas de clasificación complementarios: un máis xeral, baseado en analizar os *tokens*

en tres categorías de erro e outro con etiquetas de erro máis específicas. No primeiro sistema os erros clasifícanse en tres categorías lingüísticas: ortografía, gramática e léxico. Cada área contén tres campos de anotación: forma da palabra, lema e POS. O anotador selecciona as áreas afectadas e introduce as formas correctas.

- Nivel ortográfico: márcanse erros que producen os aprendentes na escrita, erros de forma de palabras e erros de puntuación.
- Nivel gramatical: márcanse erros que afectan á estrutura sintáctica, é dicir, que se producen cando un aprendente emprega mal unha palabra e produce enunciados agramaticais. Estes erros son os problemas de concordancia e problemas nas formas verbais.
- Nivel léxico: márcanse palabras que son ortográfica e gramaticalmente correctas pero que non empregaría un falante nativo nese contexto, é dicir, son erros que afectan ao significado.

Entre os tres niveis de análise existe unha conexión, isto é:

There is inheritance between layers, from the bottom (orthographic data) to the top (lexical data), and only what is different from the previous level has to be annotated [...] he annotator only needs to annotate what is different, and not all the fields at each layer (Río e Mendes, 2018: 228).

O sistema de clasificación con etiquetas específicas de erro (Río e Mendes, 2018: 230-231) complementa a anotación dos erros xerais nos *tokens*. As etiquetas fan posible unha clasificación de erros moito máis específica que permite facer consultas lingüísticas concretas como, por exemplo, sobre a concordancia. Para o deseño destas etiquetas neste corpus seguen uns principios teóricos:

- Deben ser o máis xerais posibles, xa que se poden crear sub-etiquetas moito máis específicas se é necesario.
- Son o máis neutrais posibles en canto ao esquema teórico para que investigadores con diferentes fondos teóricos accedan a elas facilmente.
- Ven necesario a elaboración dunha proba de anotación para desenvolver unha taxonomía de erro piloto. Esta taxonomía piloto contén 37 etiquetas.

O ASK corpus é un corpus de aprendentes de noruegués como segunda lingua. Os textos que o compoñen son composicións escritas que proveñen de dous tests diferentes de dous niveis de noruegués. O criterio base para escoller estes textos é a lingua nai dos aprendentes: alemán, holandés, inglés, español, ruso, polaco, bosnio-croata-serbio, albanés, vietnamita e somalí. Ademais, o corpus contén un corpus de control que recolle textos escritos de nativos de noruegués para comparar. O obxectivo deste corpus é servir como ferramenta de investigación na adquisición de segundas linguas (Tenjford *et al.*, 2006: 1821).

Este corpus emprega o temo *error codes* para referirse ás diferenzas que se dan entre o aprendente de noruegués e a norma. Presenta cinco categorías de erro baseadas nestas diferenzas. As categorías son (Tenjford *et al.*, 2006: 1822):

- Lexical codes: erros referidos ao léxico
- Morphological codes: error referidos á morfoloxía
- Syntactical codes: erros que afectan á sintaxe
- Punctuation codes: erros de puntuación
- Unidentified error: erros nos que non se pode identificar a intención dos autores

MERLIN, *Multilingual Platform of European Reference Levels: Interlanguage Exploration in Context*, é un corpus que explora producións escritas de aprendentes de italiano, checo e alemán.

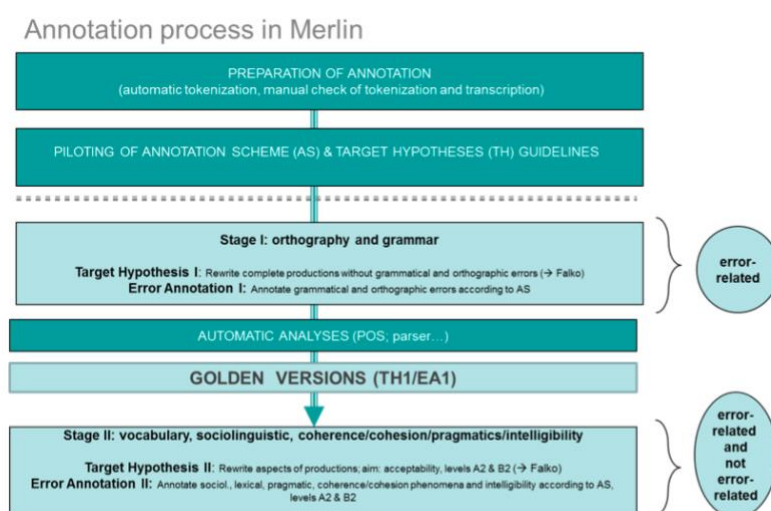


Ilustración 2. Niveis de anotación en MERLIN. Fonte: MERLIN Project (2014: 12)

Tal e como se pode ver na ilustración anterior, este sistema de anotación ten dúas fases. Na primeira, propónse unha hipótese (*Target Hypothesis I*) do que o aprendiz quere escribir para os erros ortográficos e gramaticais e anótanse este tipo de erros (*Error Anotation I*). Na segunda, ampliáanse esas hipóteses (*Target Hypothesis II*) a aspectos sociolingüísticos, léxicos e pragmáticos tendo en conta o que se espera que escriba un falante nativo e etiquétanse estes campos (*Error Anotation II*). Ademais, créanse dous niveis para mostrar as diferenzas entre o texto do aprendiz e as hipóteses que propón o anotador. Estas hipóteses non forman parte da anotación de erros pero son un requisito (MERLIN project, 2014: 15). Na seguinte táboa observamos os tipos de erros gramaticais e ortográficos que se anotan no primeiro nivel (*Error Anotation I*):

Grammar
Wrong word order (main clause/sub clause)
Errors in parts of speech
Missing/redundant obligatory complements
Errors in inflective forms (nouns, adjectives, simple and complex verbs (tense, aspect, mood, voice, use of auxiliary etc.))
Erroneous morphological aspects (case, number, gender)
Omission/addition of verbs
Errors in negation
Errors in reflexivity
Errors in the use of prepositions
Errors in the use of conjunctions
Errors in the use of clitics (ITA and CZE)
Orthography
Errors in the use of graphemes, accents and apostrophes
Erroneous abbreviations
Errors in punctuation
Errors in word splitting and merging

Ilustración 3. Erros gramaticais e ortográficos en MERLIN. Fonte: MERLIN project (2014: 28)

2.4 Sistemas de clasificación dos problemas léxicos

Atendendo aos mesmos corpus do apartado anterior, imos repasar como son os seus sistemas de clasificación de problemas considerados como léxicos dentro de cada corpus. Tamén veremos se establecen algunha división entre os problemas léxicos e semánticos.

O corpus KoKo, facendo un resumo do exposto por Abel *et al.* (2016: 3-4), ten un sistema de clasificación dos trazos léxicos que ten en conta a ocorrencia de fenómenos

léxicos así como particularidades das linguas de aprendizaxe, atendendo a dúas dimensións: á subcategoría lingüística (colocación, locución...) e á modificación con respecto ao estándar (omisión, adición etc.). O esquema que eles presentan para a anotación léxica contén 77 etiquetas. Nun procedemento de anotación multi-nivel anotan todas as ocorrencias dun fenómeno en palabras simples ou secuencias. Despois, engaden anotacións para diferenciar entre erros de corrección, erros de uso, modificacións non defectuosas –consideran aquí os usos creativos da linguaxe– e marcación diasistemática.

Para a clasificación dos erros considerados léxicos en KoKo propoñen catro categorías principais que tamén conteñen subcategorías:

1. Dimensión semántica: erros denotativos relativos á corrección ou ao uso adecuado (referencia/función, adecuación ao contexto, compatibilidade semántica e precisión) e marcas connotativas (actitude do falante e marcas diatópicas)
2. Dimensión estilística: repetición e redundancia
3. Dimensión formal: erros na formación de palabras (nivel da palabra) e erros de omisión, elección, posición e adición como modificacións creativas (nivel da frase)
4. Marcas metalingüísticas: adecuación do emprego das comiñas para destacar algún elemento

O CUTE, *Corpus of Portuguese Undergraduates' Texts*, non presenta un sistema de clasificación específico dos problemas léxicos. Dentro da súa clasificación de erros está o nivel da semántica, onde cabería esperar algunha referencia a problemas léxico-semánticos, mais este “centra-se em problemas que afetam a coesão e a coerência textuais” (Cardoso *et al.*, 2014: 176).

O *Cambridge Learner Corpus* tampouco ten unha clasificación específica de erros léxicos. Dentro da súa clasificación podemos considerar erros léxico-semánticos: *false friend errors*, *idiom errors* e *collocation errors*.

O COPLE2 na súa clasificación de erros marca no nivel léxico erros que afectan ao significado da palabra, “the operations allowed at this layer affect mainly meaning. The

word used by the learner is orthographically and grammatically correct, but it is not the most natural choice for a native speaker” (Río e Mendes, 2018: 230).

O ASK corpus recolle na súa clasificación de cinco categorías de erro unha específica de léxico. Esta categoría sinala os seguintes tipos de erro:

Lexical codes:

W	wrong word
ORT	orthographic error
PART	overcompounding
SPL	oversplitting
DER	deviant derivational affix used
CAP	deviant letter case (upper/lower)
FL	Non-Norwegian word

Ilustración 4. Erros léxicos no ASK corpus. Fonte: Tenfjord et al. (2006: 1822)

Como vemos na ilustración, non hai ningunha clasificación específica para os erros semánticos. Os erros que se encadran baixo os *lexical codes* son:

- Wrong word: selección inadecuada dunha palabra.
- Orthographic error: unidade léxica con erros ortográficos.
- Overcompounding: escritura conxunta de dúas ou máis palabras que deberían aparecer grafadas separadamente para asegurar unha ortografía correcta.
- Oversplitting: dúas formas que aparecen por separado cando deberían aparecer escritas ligadas.
- Deviant derivational affix used: afixo derivativo mal empregado.
- Deviant letter case: mal uso de maiúsculas e minúsculas.
- Non-Norwegian word: palabra que non pertence ao noruegués .

O proxecto MERLIN anota os erros léxicos na segunda fase do proceso. De acordo con Wisniewski *et al.* (2014: 18-19) estes erros divídense en:

- Xeral (*general*)
 - o erros léxico-gramaticais que levan a expresións incomprensibles: acumulación de erros léxico-gramaticais dentro dunha secuencia de palabras cando non é posible especificar os erros de xeito individual.

- formas inexistentes: palabras que non existen, calcos doutras linguas e uso de palabras ou fórmulas doutras linguas.
- erros de forma que levan a fórmulas incomprensibles: fórmulas que conteñen varios erros que fan imposible reconstruílas. A fórmula é semanticamente correcta pero a cantidade de erros na forma obstaculizan a súa comprensión.
- Erros semánticos (*semantic error*)
 - denotación: unha palabra ou fórmula emprégase mal semanticamente e fai incomprensible a proposición que a contén.
 - connotación: a palabra ou fórmula non ten un significado apropiado ao contexto. Reflicte a actitude emocional do orador/escritor cara á denotación.
 - precisión: eleccións léxicas que conducen a proposicións imprecisas semanticamente, non habituais e non nativas.
- Erros de forma (*form error*)
 - formación da palabra: a formación da palabra desvíase da convencional, erros de derivación e composición.
 - fórmulas: as fórmulas ou secuencias non son correctas respecto aos seus elementos formais (falta dunha ou máis palabras, posición incorrecta dalgunha palabra...).

3 O Corpus de textos galegos escritos por estudantes no ámbito académico (CORTEGAL)

O *Corpus de textos escritos por estudantes no ámbito académico* (CORTEGAL) ten como finalidade “contribuír a coñecer as principais dificultades que ten o alumnado á hora de escribir na variedade estándar da lingua galega, teña esta como L1 ou como L2” (Álvarez de la Granja, 2018: 55). O obxectivo é marcar todas aquelas formas que se afastan do estándar atendendo a todos os niveis lingüísticos que presenta un texto escrito.

Os textos que compoñen este corpus son os redactados para os exames da materia Lingua e literatura galegas das probas de Avaliación do Bacharelato para o acceso á Universidade (ABAU) correspondentes ao curso 2016-2017. En concreto, os textos

corresponden á pregunta na que se lles pide aos aprendentes redactar un texto de 200-250 palabras sobre un tema. Os temas que se trataron nas probas deste curso foron: a gastronomía, o consumo e a produción, os conflitos familiares e os referentes da mocidade tal e como se pode ver na seguinte ilustración.

XUÑO 2017

Opción A (texto inicial de Fran Alonso en *Dorna* 27, 2001)

Nos últimos anos a gastronomía e a cociña acadaron moita popularidade. Redacta un texto expoñendo a túa opinión sobre este fenómeno: as súas causas, o que ten de moda pasaxeira ou de cambio cultural máis duradeiro...

Opción B (texto inicial de J. Luís Sucasas en *Vieiros*, 2009)

Redacta un texto sobre a importancia que teñen o consumo e a produción (ou o consumismo e a produtividade) no noso modo de vida actual

SETEMBRO 2017

Opción A (texto inicial de Xavier Quiroga de *Zapatillas rotas*, 2014)

Expón, de maneira argumentada, a túa opinión persoal sobre o problema que reflicte o texto e, en xeral, sobre este tipo de conflitos familiares entre pais e fillos adolescentes.

Opción B (texto inicial de Mercedes Queixas, en *Palavra Comum*, 09/10/2015)

A autora móstrase crítica co feito de que a infancia e a mocidade soñe con ser futbolista ou modelo moi maioritariamente (líña 10). Redacta un texto expoñendo de maneira argumentada o teu acordo ou desacordo co seu punto de vista

Ilustración 5. Preguntas das probas ABAU. Fonte: Álvarez de la Granja (2018: 56)

Traballar con este tipo de textos supón unha serie de vantaxes e inconvenientes. Estes inconvenientes son a escasa información que se dispón dos estudantes e a necesidade de transcribir manualmente os textos. As vantaxes son varias: dispoñer destes textos de xeito inmediato, a súa homoxeneidade, a garantía da seriedade das probas, o valor e a representatividade dos textos e a posibilidade de coñecer o proceso compositivo (Álvarez de la Granja, 2018: 57).

O corpus tenta ofrecer unha transcripción fiel ao texto. Neste proceso inclúense todos os elementos do proceso de composición, incluídas as riscaduras ou cambios, así como a división en parágrafos orixinal. A diferenza doutros corpus desbótase incluír as correccións do profesor ou profesora (Álvarez de la Granja, 2018: 58-59).

En canto á anotación, as cabeceiras dos textos indicaranos a fonte do texto, o ano de composición, a convocatoria das probas (xuño/setembro), a lingua do texto, a cualificación da pregunta, a cualificación xeral da proba, o tipo de texto, o tema do texto e a comisión delegada de cada proba. A tokenización, lematización e anotación

gramatical farase de xeito automático, con posterior revisión manual (Álvarez de la Granja, 2018: 59).

Un dos obxectivos principais deste corpus é clasificar todas aquelas formas ou secuencias que se afastan do estándar. Para isto atenderase á dimensión lingüística, anotando así seis tipos de problemas (Álvarez de la Granja, 2020):

- ortográficos: problemas formais da escrita e problemas relativos ao uso das comiñas.
- morfolóxicos: problemas de carácter formal relativos á flexión.
- léxicos: emprego de unidades léxicas que non pertencen ao estándar.
- gramaticais: problemas relativos á construción de frases e oracións.
- semántico-pragmáticos: significados ou usos non estándares ou non apropiados ao contexto.
- discursivos: problemas relativos á adecuación, á coherencia e á cohesión, incluído o emprego dos signos de puntuación.

Cada problema anotarase cun código que corresponderá á estrutura I_C_D. O I identifica un dos seis ámbitos sinalados aos que pertence o problema, o C concreta o aspecto afectado (por exemplo acentuación, plurais...) e o D describe a relación ou a diferenza entre a forma do estudante e a estándar ou describe o carácter do problema (por exemplo omisión, adición...). As etiquetas correspondentes ao ámbito do problema (I) e á descrición (D) sempre deben aparecer. O aspecto afectado (C) pode aparecer ou non, dependendo do problema. Tamén se engadirá un código E para indicar a fonte ou a causa do erro no caso dalgúns problemas ortográficos, léxicos e semánticos. Ademais, unha mesma forma pode ter que anotarse en dous niveis de análise se presenta problemas de varios ámbitos ou se un erro se pode interpretar de diferentes maneiras. De existiren dúas posibles interpretacións do mesmo problema, este pode levar dúas etiquetas diferentes no mesmo nivel, por exemplo, a forma *cual* podería ser unha importación léxica do español e unha forma dialectal (Álvarez de la Granja, 2020).

4 Consecuencias léxicas do contacto do español co galego

A análise dos textos en CORTEGAL amosa unha presenza moi elevada de unidades léxicas non estándares que derivan do contacto co español. Isto correspóndese co que sucede no sistema do galego en xeral.

Neste sentido, é sabido que o contacto do español co galego produce interferencias en ambas as linguas. Sen afondarmos en temas sociolingüísticos podemos dicir que a situación das dúas linguas fai que estas interferencias teñan máis peso no galego ao presentaren durante varios séculos un status inferior ao español. Como apunta Silva Valdivia (2013: 293):

A interacción produciuse tanto na dirección galego-castelán como na castelán-galego, pero o distinto estatus dos dous idiomas fixo que as repercusións estruturais e sociais fosen moi diferentes, e que só no caso do galego se perciba como unha cuestión de verdadeira dependencia.

O termo *interferencia* foi definido por diferentes autores desde varias perspectivas. Por exemplo, Dubert (2005: 278), atendendo ás interferencias do español no galego popular, define o termo do seguinte xeito e atribuíndolle un valor diacrónico:

son interferencias los elementos lingüísticos que cumplen estas condiciones: a) aparecen recurrentemente en los enunciados, b) tienen su origen histórico en elementos tomados de enunciados adscritos a otras lenguas y c) desde el punto de vista sincrónico, funcionan en los enunciados en que aparecen al mismo nivel lingüístico que los elementos históricamente pertenecientes a la lengua en que aparecen.

E para Silva Valdivia (1991: 32) unha interferencia “trátase da transferencia dunha estrutura (en calquera dos niveis lingüísticos) entre dúas linguas que están en contacto nunha comunidade determinada”.

Con todo, Romero e Núñez (2004: 536-537) atribúenlle outro valor ao termo, establecendo diferenzas entre *préstamo*, *calco* e *interferencia*. Definen o *préstamo* a través dunha cita de Calvet (apud. Romero e Núñez, 2004: 536) na que indica que “hai préstamo lingüístico cando unha fala A utiliza e acaba por integrar unha unidade ou rasgo lingüístico que existía previamente nunha fala B e que non a posuía”. Eles engaden que os préstamos non afectan á estrutura da lingua que os acolle. O *calco* sería

un tipo de préstamo no que se copia o significado e non a fonética, por exemplo, *rañaceos* é un calco do inglés *skyscraper*. Por último, consideran a *interferencia* como “todo elemento alleo a unha lingua dada e non integrado nela, xa que cando a integración existe estamos perante un préstamo” (Romero e Núñez, 2004: 537).

Esta situación de contacto entre o galego e o español produce interferencias ou préstamos, sobre todo, no tocante ao léxico.

Esta ocupación é antiga e xeneralizada no léxico; certamente en proporcións diversas en función do ámbito temático de que se trate, pero con moi poucos espazos impermeables a esa presión. Isto fixo que bastante vocabulario que se emprega no galego popular sexa formalmente castelán, pero que se sinta funcionalmente galego (Silva Valdivia, 2013: 298).

Este feito tamén o apunta Parga Valiña (2004: 551):

O galego, desde que entrou en contacto co castelán a partir dos últimos séculos da Idade Media sufriu un proceso de redución de ámbitos de uso. O castelán converteuse en teito lingüístico do galego e abrangueu moitos dos eidos que lle pertencían co conseguinte empobrecemento paulatino do léxico autóctono. Estes campos substraídos ó galego (a algúns ámbitos nunca chegou ata a actualidade) enchéronse de solucións substitutorias procedentes do castelán.

Rodríguez Río (2003: 420) ofrece unha definición de préstamo que abarcaría o sentido global de interferencia. Considera os préstamos como unidades ou trazos lingüísticos que “unha lingua *x* posúe debido á influencia que outra lingua *y* exerce sobre ela”. Atendendo á tipoloxía de Kiesler (apud. Rodríguez Río, 2003: 421) no tocante ás unidades léxicas, establece a seguinte tipoloxía (Rodríguez Río, 2003: 422):

- Importacións: unidades que se asentan nunha lingua mantendo inalterado ou con escasas modificacións o significante e o significado que tiñan na lingua da que proceden.
 - o Estranxeirismos: aquelas importacións que conservan o significante sen cambios e ata imitan a pronuncia (ex. *smartphone*, *chef*).
 - o Préstamos léxicos: importacións que adecúan a grafía e a pronuncia ás normas da lingua que os acolle (ex. *fútbol*, *iogur*).

- Substitucións: unidades que copian un proceso denominativo (significado) empregado noutra lingua con palabras que xa existen na lingua de chegada.
 - o Calcos: combinación entre dous elementos dunha lingua copiando as pautas da formación dunha palabra noutra lingua (ex. *pasodobre*, *conlevar*).
 - o Préstamos semánticos: significantes existentes na lingua receptora que engaden un significado doutra lingua (ex. *rato*, *pobo*).
- Substitucións parciais ou préstamos híbridos: unidades nas que só un dos elementos se transferiu por importación (ex. *voz en off*).

Nesta situación de contacto do español co galego e tendo en mente a súa repercusión no léxico tamén debemos falar dos *hipergaleguismos*, denominados tamén *hiperenxebrismos*. En determinadas etapas da lingua os hipergaleguismos ligáronse a un afán diferencialista, como na segunda metade do século XIX (Fernández e Monteagudo, 1995: 123). Para Fernández e Monteagudo (1995: 124) este proceso de hipergaleguización substituiría formas lexitimamente galegas por outras ás que se lle aplicaría correspondencias contrastivas co español, por exemplo esp. *madera*-gal. *madeira* > gal. **primaveira*.

Aínda así, detrás dun hipergaleguismo non sempre está o afán diferencialista. Neste senso, Dubert (2004: 238-239) sinala que un individuo como axente activo no proceso de adquisición/aprendizaxe dunha variedade, como é o estándar, constrúe a súa lingua creando hipóteses sobre os elementos lingüísticos. A persoa establece un padrón de equivalencias comparando as formas base do español ou do galego popular coas formas do galego estándar. Así, un exemplo de hipergaleguismo sería a forma *notablemente* que un estudante nun texto non empregaría por afán diferencialista senón por considerala unha forma correcta.

5 Metodoloxía de traballo

5.1 Selección dos textos

Os textos seleccionados para este traballo foron os 300 peor cualificados dos 1000 que compoñen o corpus CORTEGAL. Estes textos son os que obtiveron peor nota na pregunta 3 –redactar un texto de 200-250 palabras sobre un tema dado– das probas da ABAU 2016-2017. A selección destes textos veu terminada pola hipótese da presenza

de máis problemas léxicos nos textos peor cualificados. A peor nota destes textos é un 0 e a máis alta un 5,83, sendo a nota media dos 300 textos un 4,42 calculada sobre 10.

5.2 Selección dos problemas

Unha vez seleccionados os textos, revisámoslos un a un e anotamos nun Excel cada problema léxico e semántico que atopabamos, incluídas as repeticións dun mesmo erro no texto. Consideramos problemas léxicos todas aquelas unidades ou expresións que se afastan da norma estándar, por exemplo as importación léxicas do español como *hacia* ou *enano*, e incluímos tamén problemas que afectan á forma de unidades gramaticais como artigos ou conxuncións, non só ás unidades léxicas propiamente ditas (por exemplo o artigo *el* en vez de *o*). Baixo os problemas semánticos recolleemos aqueles significantes estándares empregados con usos ou significados non estándares, como *desenrolar* (por *desenvolver*) ou *consumición* en *cando hai un alto nivel de **consumición** a produción aumenta*.

En moitas das clasificacións vistas ao longo deste traballo en diferentes corpus de aprendentes (vid. 2.4) non observamos unha regularidade no establecemento dos tipos de erros baixo a categoría léxico-semántica. Nós decidimos excluír deste tipo de problemas:

1. Os problemas ortográficos:
 - a. supresión/adición/substitución de grafías que non alteran a pronuncia
 - (1) *As productoras o único que buscan é **opter** o máximo beneficio*
 - b. escritura conxunta/separada
 - (2) *porque* (por *por que*)
 - (3) *demáis* (por *de máis*)
 - c. uso de maiúsculas e minúsculas
 - (4) *Master chef* (por *MasterChef*)
 - d. formas que representan unha pronuncia non estándar
 - (5) *dependendo do **contesto** social*
 - (6) *deixar de **lao** moitas cousas*

Os casos en que en español hai un grupo -cc- ou -ct- e en galego non presenta a oclusiva en posición implosiva (esp. *producto*-gal. *produto*) e as formas con

omisión dunha consoante implosiva por influencia do español cando no cognado español non existe poden considerarse erros ortográficos e léxicos. Con independencia de que tamén se poidan etiquetar como erros ortográficos, nós tratarémolos nos problemas léxicos como importacións do español:

(7) *poden facer que unha familia teña **conflictos***

(8) *fai que os fillos **sustituan** o afecto paternal ou maternal por bens materiais*

2. A adecuación dos signos de puntuación
3. As formas que se poden considerar lapsus ou equivocacións involuntarias, por exemplo:

(9) *o poder de **cambia** esta situación*

(10) *todos os seres **visvo** deste mundo*

4. Algúns problemas na conxugación verbal:

- a. alternancia vocálica

(11) *a xuventude actual **persigue** o fácil, o cómodo, o chamativo*

- b. tempos compostos

(12) *Moitos **hemos pasado** por esta fase*

- c. morfemas modotemporais

(13) ***produciendo** barato e **vendiendo** caro*

Si incluímos como problemas léxicos os problemas de flexión que afectan á raíz dos verbos irregulares e os casos nos que se emprega unha vogal temática non estándar porque o verbo se asigna a unha conxugación diferente:

(14) *non sabes nada por que **eres** un enano*

(15) *non podes **convertirte** nalgo ou nalguen polo que ves na tele*

5. Os estranxeirismos non necesarios marcados con comiñas. Consideramos que o aprendiz é consciente de que está a empregar unha forma non estándar e polo tanto non sería un problema. Por exemplo:

(16) *algún ten tamén unha sección "**Junior**"*

5.3 Clasificación establecida

Despois de observar os problemas léxicos e semánticos, ofrecemos a seguinte tipoloxía de formas non estándares cunha posibilidade de etiquetaxe partindo da estrutura I_(C)_D_(E) de CORTEGAL (vid. 3).

- Problemas léxicos
 - *importación léxica do español*: cando se dá a transferencia dunha palabra procedente do español.
 - i. *non adaptada*: se non se adapta nin á grafía, nin á fonética nin á morfoloxía (ex. *silla*). Etiqueta: Léxico_substitución_importación (L_subs_imp)
 - ii. *adaptada*: se se adecúa á grafía, fonética e/ou morfoloxía do galego (ex. *parexa*). Etiqueta: Léxico_substitución_importación adaptada (L_subs_impa)
 - iii. *híbrida*: formas que combinan unha parte importada do español con outra galega. Podemos diferenciar dous tipos: casos en que só se adapta unha parte dos compoñentes (ex. *perritos quentes*) e casos en que se mestura o galego e o español nunha única forma (ex. *inxeniería*). Etiqueta: Léxico_substitución_importación híbrida (L_subs_imph)
 - *calco*: palabras ou expresións construídas con unidades léxicas ou morfolóxicas do galego sobre esquemas morfolóxicos ou sintagmáticos do español (ex. *conlevar*). Etiqueta: Léxico_substitución_calco (L_subs_c)
 - *palabra galega desviada da norma*: son todas aquelas palabras ou expresións galegas que non se recollen no estándar, como os popularismos e os dialectalismos (ex. *millorar*). Etiqueta: Léxico_substitución_palabra galega (L_subs_pgal)
 - *creación*: palabras ou expresións inventadas con estruturas galegas (ex. *famosidade*). Etiqueta: Léxico_substitución_creación (L_subs_crea)
 - *creación por analoxía*: creación dunha palabra por analoxía con outras formas, incluímos aquí, entre outras formas posibles, os

hipergaleguismos (ex. *notablemente*). Etiqueta:
Léxico_substitución_creación analogía (L_subs_creaanal)

- *estranxeirismo non necesario*: aqueles estranxeirismos que non son necesarios por haber no galego unha palabra ou expresión que fai referencia á mesma realidade (ex. *chef*). Quedan excluídas as formas procedentes do español. Etiqueta: Léxico_substitución_estranxeirismo (L_subs_estranx)

- Problemas semánticos

- *calco semántico*: formas da lingua receptora que adquiriron novos significados por influencia do español (ex. *desenrolar*). Etiqueta: Semántico_substitución_calco semántico (S_susbs_csem)
- *palabra estándar mal empregada*: palabra do estándar que se emprega cun significado non correcto sen que a explicación se encuentre nun calco do español (ex. *actualmente a mocidade concibe máis importancia a súa roupa e ao seu físico*). Etiqueta: Semántico_substitución_palabra mal empregada (S_subs_pme)

6 Análise dos textos

A seguir, presentaremos as táboas con todos o problemas atopados. Nos problemas léxicos indicaremos o lema, as formas ou variantes dese lema –de non se indicar nada nesta columna enténdese que a forma atopada nos textos coincide co lema–, o número de informantes que empregaron esas formas, o número de veces que aparece nos textos e a forma estándar que lle corresponde a cada problema. Nos problemas semánticos indicaremos todo anterior e mais o contexto en que se atopan as formas para mostrar o valor co que se usan. Despois de cada táboa presentaremos unha breve análise dos problemas atopados. Destacaremos, entre outras cuestións, aquelas formas máis frecuentes de cada grupo.

Como xa indicamos antes, algún dos problemas poden ter varias interpretacións e por iso aparecen en varias táboas. Por exemplo, o demostrativo *eso* ou o relativo *cual* poden interpretarse como importacións do español ou como palabras galegas desviadas da norma –formas dialectais–.

6.1 Problemas léxicos

6.1.1 Importacións léxicas do español non adaptadas (L_sub_imp)

Problemas (lema)	Formas	Nº de informantes	Nº de veces que aparecen	Forma estándar
<i>a costa de</i>		1	1	á custa de
<i>a menudo</i>		1	1	a miúdo
<i>a veces</i>		2	2	ás veces
<i>abuelo</i>	<i>abuela, abuelos</i>	3	3	avoa, avó
<i>acarrear</i>		1	1	carrear
<i>acostumbrar</i>	<i>acostumbrada, acostumbrarnos, acostumbran</i>	3	3	acostumada, acostumarnos, acostuman
<i>actitud</i>		1	1	actitude
<i>asequible</i>	<i>adsequible</i>	1	1	accesible
<i>afición</i>	<i>afición, aficións</i>	4	4	afección, afeccións
<i>ahí</i>		7	7	aí
<i>ahora</i>		3	3	agora
<i>aislar</i>	<i>aislado</i>	1	1	illado
<i>alguna</i>		2	2	algunha
<i>alrededor</i>		1	1	arredor
<i>amplio</i>	<i>amplia, amplios</i>	2	2	ampla, amplos
<i>amueblar</i>	<i>amueblada</i>	1	1	amoblada
<i>antaño</i>		1	1	antano
<i>añadir</i>		1	1	engadir
<i>años</i>		1	1	anos
<i>apariencia</i>		3	3	aparencia
<i>aplastar</i>	<i>aplastando</i>	1	1	esmagando
<i>arreglar</i>	<i>arreglar, arreglamos</i>	2	2	arranxar, arranxamos
<i>arriesgar</i>	<i>arriesgado, arriesgarnos</i>	2	2	arriscado, arriscarnos
<i>artículo</i>	<i>artículo, artículos</i>	4	4	artigo, artigos
<i>aún</i>		3	3	aínda
<i>aunque</i>	<i>aun que, aunque</i>	9	10	aínda que
<i>basar</i>	<i>basar, basada, basados, basan, basándose, básase, basen</i>	10	11	basea, baseada, baseados, basean, baseándose, baséase, basean
<i>basura</i>		10	11	lixo
<i>blanco</i>	<i>blanco, blanca, blancas</i>	6	6	branco, branca, brancas
<i>bolsillo</i>	<i>bolsillo, bolsillos</i>	3	3	peto, petos
<i>bronca</i>		1	1	rifa
<i>buzón</i>		1	1	caixa do correo
<i>caiga-</i>	<i>caiga, caigamos</i>	2	2	caia, caiamos
<i>calle</i>		2	2	rúa
<i>camarero</i>	<i>camareros</i>	1	1	camareiros
<i>camino</i>		1	1	camiño
<i>carrera</i>		1	1	carreira
<i>casi</i>		1	1	case
<i>ceremonia</i>		1	1	cerimonia
<i>cierto</i>		1	1	certo
<i>ciudadano</i>	<i>ciudadanos</i>	1	1	cidadáns
<i>cocina</i>		4	4	cociña
<i>cocinar</i>	<i>cocinaba, cocinado, cocinar</i>	5	5	cociñaba, cociñado, cociñar
<i>complacer</i>	<i>complacerse</i>	1	1	compracerte
<i>comunión</i>	<i>comunións</i>	1	1	comuñóns

<i>conducta</i>		1	1	conduta
<i>conflicto</i>	<i>conflicto, conflictos</i>	13	17	conflito, conflitos
<i>conllevar</i>	<i>conlleva</i>	1	1	implica
<i>conocer</i>	<i>conoce, conocín</i>	2	2	coñece, coñecín
<i>consient-</i>	<i>consienten</i>	1	1	consenten
<i>construcción</i>		2	2	construción
<i>consuel-</i>	<i>consuela</i>	1	1	consola
<i>convertir</i>	<i>convertiuse</i>	1	1	converteuse
<i>corto</i>		2	2	curto
<i>cosa</i>	<i>cosa, cosas</i>	8	8	cousa, cousas
<i>costar</i>	<i>costa, costan, costaría</i>	2	3	custa, custan, custaría
<i>coste</i>	<i>coste, costes</i>	6	10	custo, custos
<i>costo</i>	<i>costos</i>	1	1	custos
<i>costoso</i>	<i>costoso, costosos</i>	2	2	custoso, custosos
<i>costumbre</i>	<i>costumbres</i>	1	1	costumes
<i>creciente</i>		1	1	crecente
<i>crear</i>	<i>cree, creemos, creen, creendose, crear, crees</i>	10	10	cre, cremos, cren, créndose, crer, cres
<i>crezc-</i>	<i>crezca</i>	1	1	creza
<i>crisis</i>		4	4	crise
<i>crudo</i>	<i>cruda</i>	1	1	crúa
<i>cual</i>	<i>cuales</i>	1	1	cales
<i>cumplir</i>	<i>cumprir</i>	2	2	cumprir
<i>cuota</i>		1	1	cota
<i>de-</i>	<i>de, dé, demos, den</i>	13	13	dea, deamos, dean
<i>decir</i>	<i>decir</i>	9	9	dicir
<i>dice</i>		1	1	di
<i>deconstructivista</i>		1	1	deconstrutivista
<i>delante</i>		1	1	diante
<i>demás¹</i>		1	1	de máis
<i>derrochador</i>		1	1	gastador
<i>derrochar</i>	<i>derrochamos</i>	1	1	malgastamos
<i>desaparezc-</i>	<i>desaparezca, desaparezcan</i>	2	2	desapareza, desaparezan
<i>desarrollar</i>	<i>desarrollados, desarrollar</i>	3	3	desenvolvidos, desenvolver
<i>desarrollo</i>		2	2	desenvolvemento
<i>desenfrenado</i>		1	1	desenfreado
<i>deseo</i>		1	1	desexo
<i>desgracia</i>	<i>desgracias</i>	1	1	desgrazas
<i>desigualdad</i>		1	1	desigualdade
<i>desinterés</i>		1	1	desinterese
<i>desmoronarse</i>		1	1	esborrallarse
<i>despilfarrar</i>		1	1	malgastar
<i>desprecio</i>	<i>desprecios</i>	1	1	desprezos
<i>destrucción</i>		1	2	destrución
<i>deuda</i>	<i>deudas</i>	1	1	débedas
<i>dictadura</i>		2	2	ditadura
<i>diferencia</i>	<i>diferencia, diferencias</i>	4	4	diferenza, diferenzas
<i>discurr-</i>	<i>discurrida</i>	1	1	concorrida
<i>disfrutar</i>	<i>disfruta, disfrutala, disfrutan, disfrutando, disfrutar, disfrute</i>	22	24	goza, gozala, gozan, gozando, gozar, goce
<i>disminuir</i>		1	1	diminuír
<i>doble</i>		1	1	dobre
<i>donde</i>		5	5	onde

¹ Contexto: *Por outro lado non está demás.*

<i>dueño</i>		1	1	doño
<i>dulce</i>		1	1	doce
<i>duradero</i>	<i>duradero, duradera</i>	2	2	duradeiro, duradeira
<i>duro</i>	<i>duros</i>	2	2	pesos
<i>el</i>		1	1	o
<i>ello</i> ²		1	1	isto
<i>empleado</i>		1	1	empregado
<i>empleo</i>		1	1	emprego
<i>en cuanto a</i>		1	1	en canto a
<i>en mi opinión</i>		3	3	na miña opinión
<i>en resumen</i>		1	1	en resumo
<i>enano</i>		1	1	cativo
<i>enloquecer</i>	<i>enloquece</i>	1	1	enlouquece
<i>enseñar</i>	<i>enseñar, enseñe</i>	2	2	ensinar, ensine
<i>enterarse</i>	<i>enteraste</i>	1	1	decátaste
<i>entonces</i>		1	1	entón
<i>entorno</i>		3	3	contorna
<i>entrenar</i>		1	1	adestrar
<i>entretenimiento</i>		1	1	entretemento
<i>eso</i>	<i>eso, de eso</i>	19	23	iso, diso
<i>esos</i>	<i>esos, de esos</i>	4	4	eses, deses
<i>este-</i>	<i>esté, estén</i>	16	16	estea, estean
<i>esto</i>	<i>esto, de esto</i>	39	47	isto, disto
<i>estos</i>	<i>estos, de estos, nestos</i>	21	28	estes, destes, nestes
<i>estrella</i>	<i>estrella, estrellas</i>	2	2	estrela, estrelas
<i>estricto</i>		1	1	estrito
<i>estropear</i>	<i>estropeada</i>	1	1	estragada
<i>estudiar</i>	<i>estudian, estudiar</i>	3	3	estudan, estudar
<i>estudio</i>	<i>estudio, estudios</i>	3	4	estudo, estudos
<i>eres</i>		5	5	es
<i>fastidiar</i>	<i>fastidiarnos</i>	1	1	amolarnos
<i>fichaje</i>		1	1	fichaxe
<i>filo</i>		1	1	fío
<i>finca</i>		1	1	leira
<i>frente a</i>		2	2	fronte a
<i>fruta</i>	<i>fruta, frutas</i>	2	2	froitá, froitas
<i>frutería</i>		1	1	froitaría
<i>gasolinera</i>	<i>gasolineras</i>	1	1	gasolineiras
<i>gracias</i>		2	2	grazas
<i>grado</i>	<i>grado, grados</i>	3	3	grao, graos
<i>grasa</i>	<i>grasas</i>	1	1	graxas
<i>guay</i>		2	2	estupendo
<i>hacer</i>	<i>hace, hacer, harían</i>	4	4	fai, facer, farían
<i>hacia</i>		6	6	cara a
<i>hasta</i>		7	7	ata
<i>hazaña</i>	<i>hazañas</i>	1	1	fazañas
<i>hostelería</i>		5	8	hostalaría
<i>hostelero</i>		1	1	hosteleiro
<i>hoy</i>		1	1	hoxe
<i>increíble</i>	<i>increíble, increíbles</i>	2	2	incríble, increíbles
<i>independiente</i>		1	1	independente
<i>indudable</i>		1	1	indubidable
<i>ingesta</i>		1	1	inxesta

² Contexto: Por **ello** estou de acordo con Mercedes Queixas.

<i>inmaduro</i>	<i>inmaduros</i>	1	1	inmaturós
<i>insano</i>	<i>insanas</i>	1	1	insás
<i>insostenible</i>		2	2	insostible
<i>interés</i>		1	2	interese
<i>intrascendental</i>	<i>intrascendentais</i>	1	1	intranscendentes
<i>introducción</i>		2	2	introdución
<i>interrumpir</i>	<i>a interrumpido</i>	1	1	interrompeu
<i>invertir³</i>	<i>invertir, invirten, invirtirá</i>	3	3	investir, invisten, investirán
<i>jubilación</i>		1	1	xubilación
<i>las</i>		1	1	as
<i>le</i>	<i>le, les, sumarles</i>	3	4	lle, lles
<i>leer</i>	<i>leín</i>	2	2	lín
<i>lechuga</i>		1	1	leituga
<i>legumbre</i>	<i>legumbres</i>	1	1	legumes
<i>lienzo</i>		1	1	lenzo
<i>línea</i>		1	1	liña
<i>llamativo</i>		1	1	rechamante
<i>lo cual</i>		1	1	o cal
<i>locura</i>		1	1	loucura
<i>manera</i>		1	1	maneira
<i>mano</i>		1	1	man
<i>manteng-</i>	<i>mantenga</i>	1	1	manteña
<i>más</i>		2	2	máis
<i>mayor</i>		1	1	maior
<i>mayoría</i>		1	1	maioría
<i>menospreciar</i>		1	1	menosprezar
<i>menuda⁴</i>		1	1	vaía
<i>mezcla</i>	<i>mezcla, mezclas</i>	4	4	mestura, mesturas
<i>mezclar</i>	<i>mezcla, mezclar, mezclaronse</i>	3	3	mestura, mesturar, mesturáronse
<i>mimar</i>	<i>mimados</i>	1	2	consentidos
<i>mismo</i>	<i>mismo, misma</i>	5	5	mesmo, mesma
<i>monstruo</i>	<i>monstruos</i>	1	1	monstros
<i>nadie</i>	<i>nadie, naide</i>	4	4	ninguén
<i>ni</i>		1	1	nín
<i>no</i>		4	4	non
<i>nosotros</i>		1	1	nós
<i>novedad</i>	<i>novedades</i>	2	2	novidades
<i>novedoso</i>		1	1	novo
<i>o</i>		5	5	ou
<i>obedezc-</i>	<i>obedezca</i>	1	1	obedeza
<i>ocurrir</i>	<i>ocurrir, ocurrira, ocurrirá</i>	3	3	ocorrer, ocorrerá, ocorrerá
<i>olvidar</i>	<i>olvidámonos, olvidamos, olvidándose, olvidar</i>	8	8	esquecémonos, esquecemos, esquecéndose, esquecer
<i>olvido</i>		1	1	esquecemento
<i>otro</i>	<i>otra</i>	1	1	outro
<i>padre</i>	<i>padre, padres</i>	2	2	pai, pais
<i>paladar</i>		5	5	padal
<i>pasión</i>		2	2	paixón
<i>pelear</i>	<i>pelean, peleanse</i>	2	2	pelexan, peléxanse
<i>pequeño</i>	<i>pequeña</i>	1	1	pequena

³ Contexto: *sectores importantes sobre os que se debía **invertir**.*

⁴ Contexto: *e que **menuda** presentación*

<i>pereza</i>		1	1	preguiza
<i>persona</i>	<i>persona, personas</i>	4	6	persoa, persoas
<i>personal</i>		2	2	persoal
<i>pilar</i>	<i>pilar, pilares</i>	4	4	piar
<i>pimentón</i>		1	1	pemento
<i>placer</i>		4	4	pracer
<i>plantear</i>	<i>plantearnos</i>	1	1	fórmularnos
<i>plato</i>	<i>plato, platos</i>	21	32	prato, pratos
<i>playa</i>	<i>playas</i>	1	1	praías
<i>podr-</i>	<i>podrá, podrán, podrás</i>	3	4	poderá, poderán, poderás
<i>poseer</i>	<i>posea, poseen, poseer</i>	3	4	posúa, posúe, posuír
<i>postre</i>	<i>postre, postres</i>	3	3	sobremesa, sobremesas
<i>precio</i>	<i>precio, precios</i>	14	19	prezo, prezos
<i>presencia</i>		1	1	presenza
<i>primer</i>		1	1	primeiro
<i>producción</i>		38	66	produción
<i>produzc-</i>	<i>produzca, produzcan</i>	2	2	produza, produzan
<i>productivo</i>		2	2	produtivo
<i>producto</i>	<i>produto, produtos</i>	48	96	produto
<i>productor</i>	<i>productor, productores, productoras</i>	24	35	produtor, produtoras, produtores,
<i>provechoso</i>		1	1	proveitoso
<i>proveedor</i>	<i>proveedores</i>	1	2	provedores
<i>publicidad</i>		2	2	publicidade
<i>puchero</i>		1	1	pucheiro
<i>pudiente</i>		1	1	podente
<i>pulpo</i>		4	4	polbo
<i>querr-</i>	<i>querrán</i>	1	1	quererán
<i>rabieta</i>		1	1	perrencha
<i>rato</i>	<i>rato, ratiño</i>	2	2	momento
<i>razonar</i>	<i>razonan</i>	1	1	razoan
<i>receta</i>	<i>recetas</i>	5	7	receitas
<i>rechazo</i>		1	1	rexeitamento
<i>recorrido</i>	<i>recorridos</i>	1	1	percorridos
<i>recurrente</i>	<i>recurrentes</i>	1	1	recorrentes
<i>recurrir</i>	<i>recurren, recurrirá</i>	3	3	recorren, recorrerá
<i>reducción</i>		1	1	redución
<i>reflejar</i>	<i>refleja</i>	1	1	reflicte
<i>regla</i>		1	1	regra
<i>rentable</i>	<i>rentable, rentables</i>	3	3	rendible, rendibles
<i>reponer</i>		1	1	repoñer
<i>respetable</i>		1	1	respectable
<i>respetar</i>	<i>respetadas, respetar</i>	2	2	respectadas, respectar
<i>respeto</i>		2	3	respecto
<i>resumen</i>		2	2	resumo
<i>riesgo</i>		2	2	risco
<i>riña</i>		1	1	rifa
<i>ropa</i>		4	4	roupa
<i>sabroso</i>	<i>sabroso, sabrosas</i>	2	2	saboroso, sabrosas
<i>salir</i>	<i>saldrá, sale, salen, salgamos, salir</i>	8	8	sairá, sae, saen, saíamos, saír
<i>salida</i>		2	3	saída
<i>saludable</i>	<i>saludable, saludables</i>	2	2	saudable, saudables
<i>sano</i>	<i>sano, sana, sanos</i>	7	9	san, sá, sans
<i>sea-</i>	<i>sea, seamos, sean</i>	13	13	sexa, sexamos, sexan
<i>según</i>		6	6	segundo
<i>sep-</i>	<i>sepamos, sepan, sepas</i>	4	4	saibamos, saiban, saiba
<i>servicio</i>	<i>servicios</i>	1	1	servizos

<i>si</i>		10	10	se
<i>siglo</i>		1	1	século
<i>sin</i>		5	5	sen
<i>sin duda</i>		3	3	sen dúbida
<i>sin embargo</i>		2	2	porén
<i>sinfin</i>		1	1	unha grande cantidade
<i>solamente</i>		1	1	soamente
<i>solo</i>	<i>solo, sólo</i>	10	10	só
<i>soler</i>	<i>sole</i>	1	1	adoita
<i>sordo</i>	<i>sordos</i>	1	1	xordos
<i>sostenible</i>		1	1	sostible
<i>sueldo</i>		1	1	soldo
<i>supondr-</i>	<i>supondrá, supondría</i>	2	2	supoñerá, supoñería
<i>sustituir</i>	<i>sustitúan</i>	1	1	substitúan
<i>tarea</i>	<i>tarea, tareas</i>	3	3	tarefa, tarefas
<i>tarta</i>	<i>tartas</i>	1	1	tortas
<i>telediario</i>	<i>telediarios</i>	1	1	telexornais
<i>televidente</i>	<i>televidentes</i>	1	1	telespectadores
<i>temprano</i>	<i>temprana</i>	1	1	temperá
<i>temprano</i>	<i>temprano</i>	2	2	cedo
<i>tenedor</i>	<i>tenedores</i>	1	1	garfos
<i>tener</i>	<i>tendrá, tendrán, tener</i>	5	5	terá, terán, ter
<i>tercermundista</i>	<i>tercermundista, tercer-</i> <i>mundista</i>	2	2	terceiromundista
<i>tercero</i>		1	1	terceiro
<i>término</i>	<i>término, términos</i>	4	4	termo, termos
<i>todavía</i>		4	4	aínda
<i>tozudez</i>		1	1	teimosía
<i>tontería</i>	<i>tontería, tonterías</i>	3	3	parvada, parvadas
<i>trasfondo</i>		1	1	fondo
<i>tutorial</i>	<i>tutoriais</i>	1	1	titoriais
<i>tuv-</i>	<i>tuvéramos, tuvo</i>	1	1	tiveramos, tivo
<i>una</i>	<i>una, nuna</i>	5	5	unha, nunha
<i>vaciar</i>		1	1	baleirar
<i>vacío</i>	<i>vacío, vacíos</i>	4	4	baleiro, baleiros
<i>valg-</i>	<i>valga</i>	2	2	valla
<i>va-</i>	<i>vamos</i>	8	9	imos
<i>ve-</i>	<i>véanse, veíase</i>	2	2	vexan, víanse
<i>venidero</i>	<i>venideras</i>	1	1	vindeiras
<i>venir</i>	<i>vendrán, venir</i>	2	2	virán, vir
<i>venta</i>	<i>venta, ventas</i>	6	6	venda, vendas
<i>verano</i>		1	1	verán
<i>verdad</i>		1	1	verdade
<i>víctima</i>		1	1	vítima
<i>y</i>		1	1	e

Táboa 1. Importacións léxicas do español non adaptadas

Nos textos atopamos un total de 297 importacións léxicas do español non adaptadas. Deténdonos nesta táboa, observamos que os problemas máis recorrentes son *producto* (96 ocorrencias en 48 informantes) e *producción* (66 ocorrencias en 38 informantes) Tamén destacamos *productor* (35 ocorrencias en 24 informantes). Como tratamos na clasificación, estes problemas cos grupos -ct- e -cc- poderían ser considerados como ortográficos pero vendo o alto número de aparición nos textos –tamén aparecen as

formas *conflicto*, *afición*, *respeto*, *construcción*, *destrucción*, *introducción*...– consideramos que se deben ao contacto co español.

Cómpre destacar tamén os problemas cos demostrativos (e as súas contraccións) *esto* (47 ocorrencias en 39 informantes), *estos* (28 ocorrencias en 21 informantes) e *eso* (23 ocorrencias en 19 informantes). Cabe a posibilidade de interpretar tamén estes problemas –só *esto* e *eso*, xa que a forma *estos* é de orixe castelá– como palabras galegas desviadas da norma e así o indicamos na táboa correspondente (vid. 6.1.5). Non podemos asegurar ao cen por cento que sexan un tipo ou outro de problemas xa que non traballamos con datos biográficos dos informantes nin con variables sociolingüísticas e aínda que si podemos coñecer a área na que realizan as probas a través das Comisións delegadas, este dato non é determinante para establecer se nos atopamos ante formas dialectais ou importacións. Con todo, coñecedores da situación actual do galego, é bastante probable que estes problemas se expliquen tamén pola influencia do español.

Outros problemas recorrentes son o substantivo *plato* (32 ocorrencias en 21 informantes) e o verbo *disfrutar* (24 ocorrencias en 22 informantes).

6.1.2 Importacións léxicas do español adaptadas (L_sub_imp)

Problemas (lemas)	Formas	Nº de informantes	Nº de veces que aparecen	Forma estándar
<i>afloxar</i>	<i>afloxa</i>	1	1	afrouxa
<i>alexar</i>	<i>alexada</i>	1	1	afastada
<i>antiguedade</i>	<i>antiguedade, antiqüedade</i>	2	2	antigüidade
<i>antoxo</i>	<i>antoxos</i>	1	1	antollos
<i>artiluxio</i>	<i>artiluxios</i>	3	3	dispositivos
<i>aquelo</i>		2	3	aquilo
<i>atraz-</i>	<i>atrazo</i>	1	1	atraeu
<i>calquer</i>		11	12	calquera
<i>correxir</i>	<i>correxidos</i>	1	1	corrixidos
<i>cotidián</i>	<i>cotidiá, cotidián</i>	3	3	cotiá, cotián
<i>cualidade</i>		1	1	calidade
<i>desvanecemento</i>		1	1	esvaecemento
<i>dineiro</i>		1	1	diñeiro
<i>elixir</i>		3	4	elixir
<i>elo</i>		4	4	iso
<i>entreteñemento</i>		2	2	entretemento
<i>esclavitude</i>		1	2	escravitude
<i>estreia</i>	<i>estreias</i>	1	1	estrelas
<i>expoñ-</i>	<i>expoñe</i>	1	1	expoñe
<i>fac-</i>	<i>facen</i>	1	1	fan
<i>financieiro</i>	<i>financieiras</i>	1	1	financeiras
<i>ganadeiro</i>	<i>ganadeiros</i>	1	1	gandeiros

<i>hai-</i>	<i>haia</i>	2	2	<i>haxa</i>
<i>hosteleiro</i>	<i>hosteleiros</i>	1	1	<i>hostaleiros</i>
<i>incapé</i>		1	1	<i>fíncapé</i>
<i>incertidume</i>		1	1	<i>incerteza</i>
<i>introdux-</i>	<i>introduxo</i>	1	1	<i>introduciu</i>
<i>inxenieiro</i>		1	1	<i>enxeñeiro</i>
<i>manteñ-</i>	<i>manteñemos</i>	1	1	<i>mantemos</i>
<i>marxen</i>	<i>márxenes</i>	1	1	<i>marxes</i>
<i>mitade</i>		1	1	<i>metade</i>
<i>mortandade</i>		1	1	<i>mortalidade</i>
<i>obxetivo</i>	<i>obxetiva, obxetivo</i>	2	2	<i>obxectiva, obxectivo</i>
<i>obxeto</i>	<i>obxetos</i>	4	4	<i>obxectos</i>
<i>perxudicar</i>	<i>perxudica, perxudican, perxudicar, perxudicarnos, perxudique</i>	4	6	<i>prexudica, prexudican, prexudicar, prexudicarnos, prexudique</i>
<i>perxudicial</i>	<i>perxudicial, perxudiciais</i>	2	2	<i>prexudicial</i>
<i>placenteiro</i>		1	1	<i>pracenteiro</i>
<i>plantexar</i>	<i>plantexar, plantexarnos, plantéxanse</i>	3	3	<i>formular, formularnos, formúlanse</i>
<i>pu-dr-s</i>	<i>podren</i>	1	1	<i>podrecen</i>
<i>productividade</i>		4	5	<i>produtividade</i>
<i>produx-</i>	<i>produxo</i>	1	1	<i>produciu</i>
<i>razonamento</i>	<i>razonamentos</i>	1	1	<i>razoamentos</i>
<i>reflexar</i>	<i>reflexa reflexado, reflexan</i>	3	3	<i>reflicte, reflectido, reflecten</i>
<i>soer</i>	<i>soe, soen</i>	7	7	<i>adoita, adoitan</i>
<i>sosteñ-</i>	<i>sosteñerse</i>	1	1	<i>sosterse</i>
<i>surxir</i>	<i>surxiron</i>	1	1	<i>xurdiron</i>
<i>teñ-</i>	<i>teñe, teñer</i>	2	3	<i>ten, ter</i>
<i>terneira</i>		1	1	<i>tenreira</i>
<i>trabago</i>		1	1	<i>traballo</i>
<i>trampantoxo</i>	<i>trampantoxos</i>	1	1	<i>trampantollos</i>
<i>ventaxa</i>	<i>ventaxas</i>	2	2	<i>vantaxes</i>
<i>xeneralmente</i>		1	1	<i>xeralmente</i>
<i>xenerando</i>		1	1	<i>xerando</i>
<i>xolla</i>	<i>xollas</i>	1	1	<i>xoias</i>
<i>xoven</i>	<i>xoven, xóvenes</i>	14	16	<i>mozo, mozos</i>
<i>xuez</i>	<i>xueces</i>	1	1	<i>xuíces</i>

Táboa 2. Importacións léxicas do español adaptadas

Das 56 importacións léxicas adaptadas atopadas nos textos, destacamos o substantivo *xoven* (16 ocorrencias en 14 informantes) e o identificador *calquer* (12 ocorrencias en 11 textos) como as máis recorrentes. Tamén é interesante a adaptación de formas como *teñer* e derivados (*entreteñemento, manter e sosteñerse*).

Analizando a táboa, podemos establecer os procedementos máis frecuentes que empregan os aprendentes para adaptar as formas:

5 Interpretamos como forma base o infinitivo *pu-drir* que se conxugaría como *durmir*.

- adaptación ao galego de sufixos e terminacións:
 - o -dad / -tad > -dade / -tade: *antigüedad, metade, mortandade, productividade*
 - o -tud > -tude: *esclavitude*
 - o -ero > -eiro / -era > -eira: *ganadeiro, hosteleiro, inxenheiro, placenteiro, dineiro, financeiro, placenteiro, terneira*
 - o -ano > -án: *cotidián*
- substitución das grafías *g, j* por *x*: *afloxar, alexar, antoxo, artiluxio, atraxo, correxir, elixir, introduxo, marxén...*
- despalatalización da lateral: *aquelo, elo*
- monotongación dos ditongos crecentes *-ua-* e *-ie-* : *calquer, incapé, razonamento, desvanecemento*
- palatalización da nasal alveolar: *teñer* e derivados

6.1.3 Importacións léxicas do español híbridas (L_subsub_imph)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>cumpleanos</i>		2	2	aniversario
<i>desvantaxa</i>	<i>desvantaxas</i>	2	3	desvantaxes
<i>inxeniería</i>		1	1	enxeñería
<i>perrito quente</i>	<i>perritos quentes</i>	1	1	-----
<i>pescadilla que se morde a cola</i>	<i>pescadilla que mórtese a cola</i>	1	1	a carioca co rabo na boca
<i>picar o gusaniño</i>	<i>picoulle o gusaniño</i>	1	1	entrar o bicho
<i>vantaxa</i>	<i>vantaxas</i>	1	2	vantaxes

Táboa 3. Importacións léxicas do español híbridas

A forma máis recorrente das importacións léxicas do español híbridas é *vantaxa* que xunto con *desvantaxa* teñen 5 ocorrencias en 3 informantes. Esta forma créase ao xuntar o galego *vantaxe* co español *ventaja*.

Como xa indicamos na clasificación, temos dous tipos de formas híbridas que se reflicten nesta táboa:

- casos en que só se adapta un dos compoñentes: *cumpleanos, perrito quente, pescadilla que se morde a cola* e *picar o gusaniño*
- casos en que se mestura o galego e o español: *desvantaxa, inxeñería* e *vantaxa*

6.1.4 Calcos (L_subsub_c)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>alonxar</i>		1	1	afastar
<i>calqueira</i>		7	7	calquera
<i>chamativo</i>		1	1	rechamante
<i>chear</i>	<i>chear, cheas</i>	2	2	encher, enches
<i>conlevar</i>	<i>conleva, conlevan, conlevar</i>	9	9	implica, implican, implicar
<i>cumpleanos</i>		2	2	aniversario
<i>fornear</i>	<i>forneada</i>	1	1	enforxada
<i>lonxano</i>	<i>lonxano, lonxanos</i>	2	2	afastado, afastados
<i>perrito quente</i>	<i>perritos quentes</i>	1	1	-----
<i>pescadilla que se morde a cola</i>	<i>pescadilla que mórdese a cola</i>	1	1	a carioca co rabo na boca
<i>picar o gusaniño</i>	<i>picoulle o gusaniño</i>	1	1	entrar o bicho
<i>rechea</i>		1	1	reencher
<i>sen embargo</i>		11	12	porén
<i>xeralizar</i>	<i>xeralizada</i>	1	1	xeneralizada

Táboa 4. Calcos

Dos 14 calcos presentes nos textos analizados, temos que as formas que se calcan son derivados (*alonxar, chamativo, chear, conlevar, fornear, lonxano, rechea* e *xeralizar*), un conector discursivo pluriverbal (*sen embargo*), un identificador no que se fusionan dúas palabras (*calqueira*), tres expresións idiomáticas (*perrito quente, pescadilla que se morde a cola* e *picar o gusaniño*) e un composto (*cumpleanos*).

As expresións *perrito quente, pescadilla que se morde a cola* e *picar o gusaniño* e o compostos *cumpleanos* poden ser considerados importacións híbridas, como xa vimos na táboa anterior, pero tamén se poden analizar como calcos xa que en todos os casos se calca a estrutura morfosintáctica do español.

As formas máis recorrentes desta táboa son *sen embargo* (12 ocorrencias en 6 informantes), *conlevar* (9 ocorrencias en 9 informantes) e *calqueira* (7 ocorrencias en 7 informantes). En canto a *calqueira*, recordamos que este identificador tamén é unha das formas recorrentes nas importación adaptadas coa variante *calquer*.

Sen embargo é bastante habitual tanto na fala coma na escrita. Non é de estrañar que se produza o calco xa que é un dos marcadores contraargumentativos de uso máis frecuente no español.

Ademais, tamén podemos salientar que para a forma galega *afastar* se producen dous tipos de problemas nestes textos, ou ben temos unha importación adaptada do español como é *alexar* ou ben un calco como *alonxar*, ligado tamén ao adxectivo *lonxano*.

6.1.5 Palabras galegas desviadas da norma (L_sub_pgal)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>asesmo</i>		1	1	así mesmo
<i>aquelo</i>		2	3	aquilo
<i>cual</i>	<i>cuales</i>	1	1	cales
<i>en cuanto a</i>		1	1	en canto a
<i>eso</i>	<i>eso, deso</i>	19	23	iso, diso
<i>esteñ-</i>	<i>esteñan</i>	1	1	estean
<i>esto</i>	<i>esto, de esto</i>	39	47	isto
<i>ise</i>	<i>ise, ises</i>	2	2	ese
<i>iste</i>	<i>distes</i>	1	1	destes
<i>límete</i>		1	1	límite
<i>millor</i>	<i>millor, millores</i>	2	2	mellor, mellores
<i>millorar</i>		1	1	mellorar
<i>nembargantes</i>		1	1	porén
<i>pa</i>		1	1	para
<i>pidir</i>	<i>pidindo</i>	1	1	pedindo
<i>poid-</i>	<i>poidese, poidesen</i>	2	2	puidese, puidesen
<i>pra</i>		2	2	para
<i>semá</i>	<i>semás</i>	1	1	semanas
<i>señ-</i>	<i>seña</i>	1	1	sexa
<i>tar</i>	<i>ta, tase, tou</i>	2	3	está, estase, estou

Táboa 5. Palabras galegas desviadas da norma

As formas máis recorrentes das palabras galegas derivadas da norma son os demostrativos *esto* (47 ocorrencias en 39 informantes) e *eso* (23 ocorrencias en 19 informantes). Nesta táboa tamén aparecen outras variantes dialectais dos demostrativos estándares como *aquelo* (3 ocorrencias en 2 informantes), *ise* (2 ocorrencias en 2 informantes) e *iste* (1 ocorrencia en 1 informante). Como xa vimos antes, algunhas destas formas tamén poden ser analizadas como importacións léxicas adaptadas ou non, segundo o caso (*esto, eso, aquilo*). Igualmente, as formas *cual* e *en cuanto a* tamén se poden considerar ou ben importacións léxicas do español, ou ben formas dialectais características do bloque oriental.

Dentro deste tipo de problemas atopamos sobre todo dous grupos:

- dialectalismos: *esto, eso, aquilo, tar, ise, cual, en cuanto a, esteñan, iste, poidese, semá, seña*
- popularismos: *asesmo, límete, millor, millorar, pa, pidir, pra*

Facendo unha procura rápida no *Tesouro do léxico patrimonial galego e portugués* vemos que a variante *semá* aparece recollida no Valadouro e en Castroverde polo que tamén decidimos incluír esta forma como un dialectalismo. Tamén é certo, como veremos, que se pode ver como un hipergaleguismo.

A conxunción adversativa *nembargantes* era considerada forma normativa no 1982, nas propias normas apuntan a unha distribución xeográfica específica:

No galego actual non hai testemuños senón de *sen embargo*, *sin embargo* e *nembargante(s)*, esta última forma de ámbito xeográfico moi restrinxido. [...] *Nembargante(s)* está vivo aínda hoxe, e, a pesar de non existir case en castelán e non existir en portugués, é unha forma lexítima no galego (NOMIG, 1995: 184).

No ano 2003 esta forma foi desbotada por considerarse interferencia do español. Así o explica Freixeiro (2005: 256):

Nembargante(s) ou *no(n) embargante(s)* pode considerarse unha variante de *sen embargo* pola base lexical común (*embargo*) e por partillaren usos discursivos [...] o relativamente importante uso de *sin embargo* (absolutamente dominante inicialmente) ou *sen embargo* no galego escrito apuntan na vía de aproximación ao castelán e, por tanto, fan sospeitar da posibilidade de interferencia desta lingua, con cuxos usos e valores pragmáticos coinciden neste caso.

Con todo, foi moi empregada na escrita con afán diferencialista e leva asociada, aínda hoxe, a marca de forma culta.

6.1.6 Creacións (L_sub_crea)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>asombrante</i>		1	1	asombroso
<i>ceda</i>		1	1	temperá
<i>clientera</i>		1	1	clientela
<i>divulgamento</i>		1	1	divulgación
<i>en concludinte</i>		1	1	en conclusión
<i>en conseguinte</i>		1	1	por conseguinte
<i>famosidade</i>		1	1	fama
<i>produtivismo</i>		1	1	produtividade

Táboa 6. Creacións

Todas as creacións teñen só unha ocorrencia nos textos e polo tanto son casos illados. Nalgunhas os aprendentes empregan sufixos derivativos frecuentes no galego como en

asombrante, divulgamento, famosidade e produtividade ou propios do español, como no caso de *clientera* (aínda que esta forma se podería considerar tamén un lapsus). Noutras crean conectores de constitución sintagmática como *en concludinte* e *en conseguinte*.

É interesante a creación de *ceda* como adxectivo feminino. O aprendente toma o adverbio de tempo *cedo* e flexiona ao feminino tendo así:

Normalmente, a ceda idade dos 15 - 16 anos

6.1.7 Creacións por analoxía (L_sub_creaanal)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>a pesares</i>		1	1	a pesar
<i>acesibre</i>	<i>acesibres</i>	1	1	acesibles
<i>aprezado</i>	<i>aprezada</i>	1	1	apreciada
<i>compricado</i>		1	1	complicado
<i>concenciar</i>		1	1	concienciar
<i>consomer</i>		1	1	consumir
<i>convivenza</i>		1	1	convivencia
<i>gañanza</i>	<i>gañanza, gañanzas</i>	2	2	ganancia, ganancias
<i>inconvenente</i>	<i>inconvenentes</i>	1	1	inconvenientes
<i>inconvinte</i>		1	1	inconveniente
<i>influenza</i>		1	1	influenza
<i>isa</i>		2	3	esa
<i>ista</i>		4	4	esta
<i>istos</i>	<i>distos</i>	1	1	destes
<i>meiro</i>		1	1	mero
<i>notabremente</i>		1	1	notablemente
<i>nutrinte</i>	<i>nutrintes</i>	1	1	nutrientes
<i>oficiña</i>	<i>oficiñas</i>	1	1	oficinas
<i>persoaxe</i>	<i>persoaxes</i>	5	6	personaxes
<i>pod-</i>	<i>podeu</i>	1	1	puido
<i>pranta</i>	<i>prantas</i>	1	1	plantas
<i>puid-</i>	<i>puidendo</i>	1	1	podendo
<i>semá</i>	<i>semás</i>	1	1	semanas
<i>sinceiro</i>	<i>sinceira</i>	1	1	sincera
<i>unhos</i>		5	7	uns
<i>vacina</i>		1	1	vacina

Táboa 7. Creacións por analoxía

Nos textos analizados atopamos 25 creacións por analoxía. As máis recorrentes son *unhos* (7 ocorrencias en 5 informantes) e *persoaxe* (6 ocorrencias en 5 informantes).

Dentro destas creacións por analoxía podemos distinguir tres grupos:

- as formas analóxicas: a partir doutras formas do seu paradigma. Temos: *unhos, ista, ista, istos, puidendo, consomer* e *a pesares*.

- os hipergaleguismos: estes créanse aplicando patróns de equivalencia entre parellas de formas galegas e españolas. Nos desta táboa vemos os seguintes procedementos:
 - o muda da líquida lateral *l* pola líquida vibrante *r* seguindo o patrón de esp. *blando* – gal. *brando*: *accesibre, complicado, notablemente, pranta*
 - o muda do sufixo *-cia* por *-za* seguindo o patrón de esp. *diferencia* – gal. *diferenza*: *convivenza, apreciado, gañanza, influenza*
 - o muda da terminación *-ana* por *-á* seguindo o patrón de esp. *manzana* – gal. *mazá*: *semá*
 - o monotongación do ditongo crecente *-ie-* seguindo o patrón de esp. *siguiente* – gal. *seguinte*: *concenciar, inconveniente*
 - o palatalización da nasal alveolar seguindo o patrón de esp. *cocina* – gal. *cociña*: *gañanza, oficiña, vaciña*
 - o caída do *n* intervocálico seguindo o patrón de esp. *persona* –gal. *persoa*: *persoaxe*
 - o ditongación de *e* seguindo o patrón de esp. *caldero* – gal. *caldeiro*: *meiro, sinceiro*
- as formas analóxicas a partir doutras formas galegas: non se aplica un patrón de equivalencia do español senón que se regulariza a gramática. Estas son: *inconvinte, nutritante e podeu*.

Como indica Dubert (2004: 241) nun artigo en que trata formas do tipo de *inconvinte* e *nutrinte*, “as formas non-estándar créanse en virtude dunha tendencia a regularizar por medio da analoxía a formación destes adxectivos en *-nte*”. Neste caso prodúcese a extensión de *-i* a lugares en que, na variedade estándar, debora aparecer *-ie-* (cunha regularización a maiores na raíz no caso de *inconvinte*).

6.1.8 Estranxeirismos non necesarios (L_subst_estrax)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar
<i>boom</i>		4	5	auxe
<i>celebrity</i>	<i>celebrities</i>	1	1	famosos
<i>chef</i>	<i>chef, chefs</i>	12	13	xefe de cociña
<i>detox</i>		1	1	desintoxicante

<i>exprés</i>		1	1	rápidas
<i>spam</i>		1	1	anuncios lixo

Táboa 8. Estranxeirismos non necesarios

Como xa indicamos, para a clasificación deste tipo de problemas desbotamos todos aqueles estranxeirismos non necesarios que os aprendentes escriben entre comiñas, entendendo que o propio estudante é consciente de que está a empregar unha forma non estándar, por exemplo:

a que tomas para ser o máis "cool"

Así, os estranxeirismos máis recorrentes nestes textos son *chef* (13 ocorrencias en 12 informantes) e *boom* (5 ocorrencias en 4 informantes). A forma *chef* provén do francés e a forma *boom* do inglés. De todas as formas desta táboa, *chef* tamén aparece entre comiñas nalgún texto:

converte aos "chefs" en persoaxes famosas

6.2 Problemas semánticos

6.2.1 Calcos semánticos (S_sub_csem)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar	Contexto
<i>bacharel</i>		1	1	bacharelato	<i>grados medios e superiores que substitúen ao bacharel</i>
<i>canal</i>	<i>canais</i>	1	1	canles	<i>hai canais dedicados exclusivamente á cociña</i>
<i>desenrolar</i>	<i>desenrolado, desenrolar</i>	4	4	desenvolvido, desenvolver	<i>este ámbito de vida estaba moi pouco desenrolado</i>
<i>desenrolo</i>		1	1	desenvolvemento	<i>desenrolo cultural</i>
<i>facer</i>	<i>fai</i>	25	25	hai	<i>fai uns anos, a xente compraba só o necesario</i>
<i>largo</i>	<i>largo, larga</i>	3	3	longo, longa	<i>o camiño corto mándache despois ao largo</i>
<i>largo</i>	<i>ao largo de</i>	1	1	ao longo de	<i>ao largo da súa vida</i>
<i>pobo</i>	<i>pobo, pobos</i>	5	5	vila, vilas	<i>hai bares nos pobos</i>
<i>sono</i>		2	2	soño	<i>sacrifícase para cumprir o seu sono</i>

Táboa 9. Calcos semánticos

Dos 9 calcos semánticos atopados nos textos hai un que é o máis recorrente con diferenza, a forma *fai*. Este calco semántico prodúcese ao empregar o verbo *facer* como

no español, concretamente a P3 do presente de indicativo (*fai*), no canto do verbo *haber* en expresións de tempo, por exemplo:

Fai uns anos, a xente compraba só o necesario (GAL)

Hace unos años la gente compraba solo lo necesario (ESP)

6.2.2 Palabras estándar mal empregadas (S_subst_pme)

Problemas (lema)	Formas	Nº de informantes	Nº de veces	Forma estándar	Contexto
<i>agora</i>		1	1	á hora	<i>sintense indefensos agora de deixarnos</i>
<i>alimenticio</i>	<i>alimenticios</i>	1	1	alimentarios	<i>no caso de produtos alimenticios</i>
<i>ámbito</i>	<i>ámbitos</i>	2	2	hábitos	<i>este é un cambio cultural dos ámbitos alimenticios que favorecerá unha maior saúde</i>
<i>comunicativo</i>	<i>comunicativos</i>	2	2	de comunicación	<i>foi impulsado polos medios comunicativos</i>
<i>con respecto a</i>		1	1	referidos á	<i>dispoñemos de moitos programas con respecto a cociña</i>
<i>concibir</i>	<i>concibe</i>	1	1	dálle	<i>actualmente a mocidade concibe máis importancia a súa roupa e ao seu físico</i>
<i>consumición</i>	<i>consumición, consumicións</i>	11	16	consumo, consumos	<i>cando hai un alto nivel de consumición a produción aumenta</i>
<i>consumismo</i>		14	16	consumo	<i>o consumismo estaría ben en certa medida, sempre e cando un saiba poñer límites</i>
<i>consumista</i>		1	1	consumidora	<i>exercer de consumista e prefire comprar algo menos elegante</i>
<i>corrosco</i>		1	1	cornete	<i>podese ver a unha persoa comendo un corrosco de croquetas</i>
<i>dar a luz</i>		1	1	sacar á luz	<i>temos todos os instrumentos e medios necesarios para dar a luz este tema</i>
<i>desquitar</i>	<i>desquita</i>	1	1	quita	<i>a xente non se desquita e priva dun bó almorzo</i>
<i>discorrer</i>	<i>discurrida</i>	1	1	concorrida	<i>Actualmente é moi discurrida entre a xente</i>
<i>evocación</i>	<i>evocacións</i>	1	1	vocacións	<i>existen diferentes tipos de xóvenes e polo tanto diferentes tipos de evocacións</i>
<i>expedición</i>		1	1	inspección	<i>demáis que lle garante a legalidade laboral no caso de producirse unha expedición</i>
<i>gastronómico</i>	<i>gastronómicos</i>	1	1	gastrónomos	<i>cada vez aparecen mais gastronómicos, ou chef</i>
<i>indixesta</i>		1	1	inxesta	<i>pola indixesta sen control de comida basura</i>

<i>instancia</i>		1	1	existencia	<i>Os seres humanos son mamíferos evolucionados dos primates, que basean a súa instancia nun réxime capitalista e consumista.</i>
<i>intre</i>		1	1	momento	<i>sen ter que estar cada intre pendentes dos seus fillos</i>
<i>logo</i>	<i>logo que</i>	1	1	xa que	<i>estudan e traballan á vez, logo que a súa carreira pode truncarse dun momento a outro</i>
<i>malo</i>	<i>facen malo</i>	1	1	facen mal	<i>entre outros que nos van facer malo para á nosa saúde</i>
<i>mineiro</i>	<i>mineiras</i>	1	1	minas	<i>sacarán o ouro das mineiras</i>
<i>onde</i>		1	1	na que	<i>nesta época onde vivimos</i>
<i>onte</i>		1	1	antes	<i>como dixera onte</i>
<i>porén</i>		1	1	por iso	<i>Non obstante, non só é necesario unha dieta variada para vivir saudablemente, tamén facer exercicio diariamente é un paso necesario. Porén, esta moda das dietas exprés que popularizan certos libros non é necesaria se un sabe coidarse.</i>
<i>promover</i>	<i>promovida</i>	1	1	motivada	<i>Así a xente promovida por estes programas comeza a cociñar</i>
<i>situar</i>	<i>sitúase</i>	1	1	está	<i>a gastronomía sitúase en auxe</i>
<i>suxestionar</i>	<i>suxestionada</i>	1	1	programada	<i>"obsolescencia suxestionada"</i>

Táboa 10. Palabras estándar mal empregadas

A maioría das palabras estándar mal empregadas atopadas nos textos teñen só unha ocorrencia, son casos illados. Destacan as formas recorrentes *consumismo* (16 ocorrencias en 14 informantes) e *consumición* (11 ocorrencias en 16 informantes). Ambas formas empréganse co valor de “consumo”. As dúas formas pertencen á mesma familia léxica.

Observando a táboa podemos establecer unha especie de tipoloxía parcial con algunhas formas:

- confusión en formas que pertencen á mesma familia léxica: *alimenticios* (por *alimentarios*) *consumismo* (por *consumo*), *consumición* (por *consumo*), *consumista* (por *consumidor*) *comunicativo* (por *de comunicación*) *gastronómico* (por *gastrónomo*) *malo* (por *mal*) e *mineiras* (por *minas*).

- confusión en formas que se parecen formalmente: *agora* (por *á hora*), *ámbito* (por *hábito*), *dar a luz* (por *sacar á luz*), *evocación* (por *vocación*), *expedición* (por *inspección*) e *indixesta* (por *inxesta*).
- confusión en formas que nalgún contexto se parecen semanticamente: *intre* (por *momento*), *situarse* (por *estar*) *desquitar* (por *quitar*) e *promovida* (por *motivada*).
- confusión de conectores: *logo que* (por *xa que*) e *porén* (por *por iso*).

7 Conclusións

Unha vez analizadas as táboas na sección anterior (vid. 6) podemos dicir que o número total de lemas é de 474 e o total de formas de 1525. A seguir ofreceremos gráficos correspondentes aos lemas e despois os correspondentes ás formas proporcionadas por informantes diferentes.

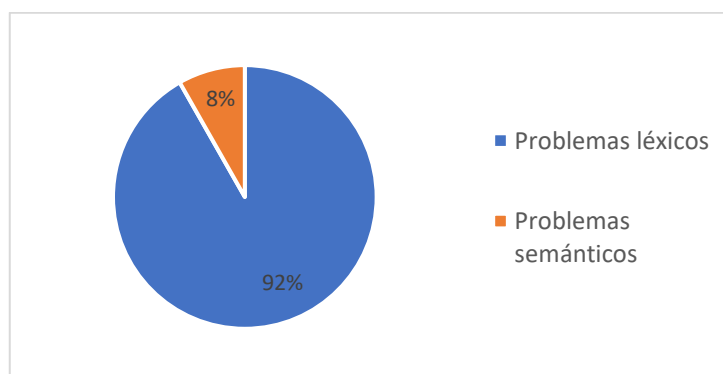


Gráfico 1. Problemas léxicos e semánticos por lemas

Atendendo aos lemas, vemos que os problemas léxicos son os máis frecuentes cun 92% (435 lemas) fronte aos semánticos cun 8% (39 lemas).

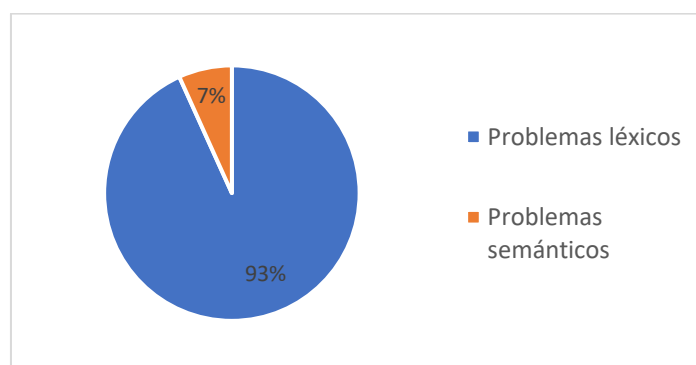


Gráfico 2. Problemas léxicos e semánticos por formas

Observando o segundo gráfico vemos que a maioría dos problemas atopados en canto ás formas son os léxicos cun 93% de formas (1422) fronte aos semánticos cun 7% de formas (103).

Dentro dos problemas léxicos os máis habituais son as importacións do español. Atendendo aos lemas temos que estas abranguen o 83% (361 lemas).

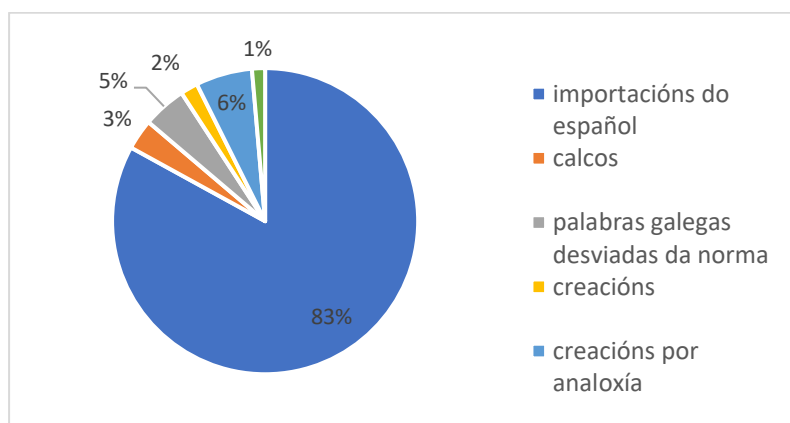


Gráfico 3. Problemas léxicos por lemas

Se prestamos atención ás formas as importacións do español seguen a ser as máis habituais cun 85% de formas (1075), seguidas das palabras galegas desviadas da norma cun 7% (96), os calcos cun 3% (42), as creacións por analogía cun 3% (43), os estranxeirismos non necesarios cun 1% (22) e as creacións cun 1% (8).

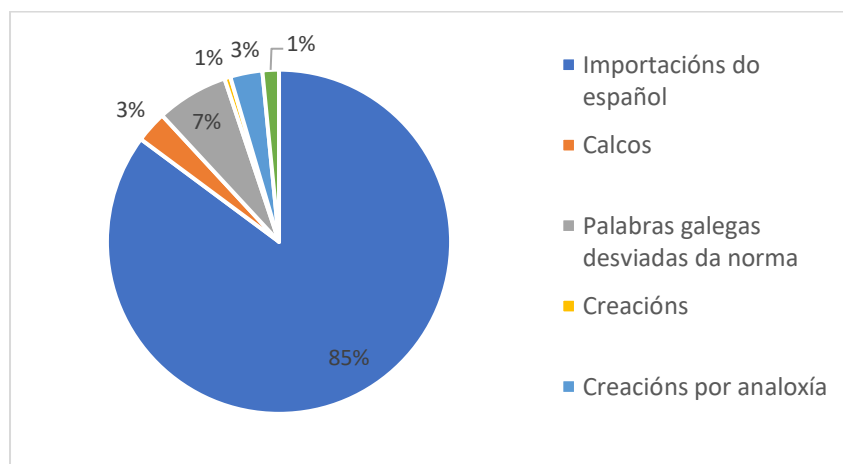


Gráfico 4. Problemas léxicos por formas

As importacións clasificábanse en tres grupos: non adaptadas, adaptadas e híbridas. Como se ve nos seguintes gráficos, as non adaptadas teñen máis peso sobre as outras

cun 83% (298) dos lemas e cun 89% das formas (1075). Aínda así, máis da metade destas importacións non adaptadas teñen por debaixo de 5 ocorrencias.

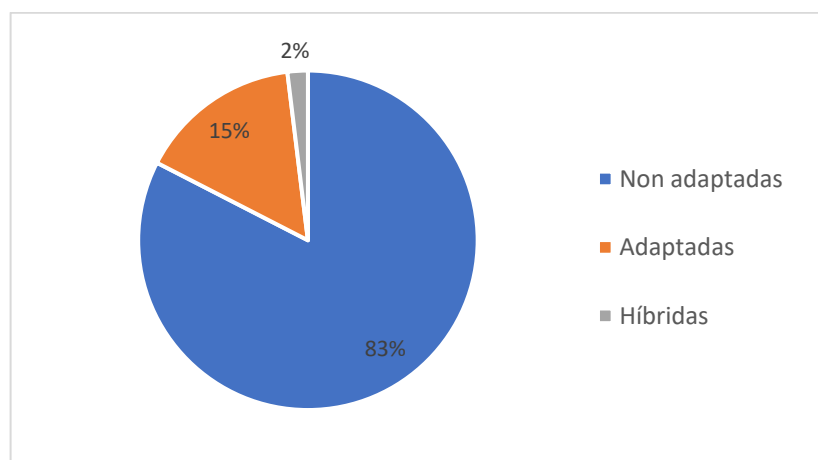


Gráfico 5. Importacións léxicas do español por lemas

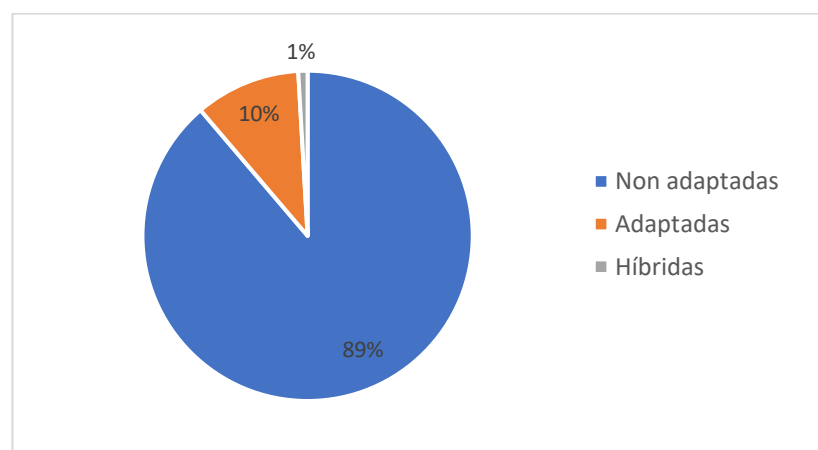


Gráfico 6. Importacións léxicas do español por formas

As formas máis recorrentes dentro das importacións non adaptadas son da familia léxica de *product* e o substantivo *plato*. Non obstante, isto non quere dicir que estas formas teñan necesariamente un maior grao de integración ca outras formas que aparecen en menor medida ou mesmo que non aparecen, xa que os textos tiñan uns temas concretos preestablecidos –a gastronomía, o consumo e a produción, os conflitos familiares e os referentes da mocidade–, o que condiciona claramente a aparición de certas formas. Neste caso *product* aparece, sobre todo, en textos sobre o consumo e a produción e *plato* en textos sobre a gastronomía.

No tocante aos problemas semánticos os máis habituais nestes textos son as palabras estándar mal empregadas cun 77% dos lemas e un 58% de formas (43). Como xa

sinalamos, a maioría destas formas son casos illados e só destacan por recorrentes *consumismo* e *consumición* cos valores de “consumo”. Estas seleccións tamén están condicionadas polo tema do texto, consumo e produción.

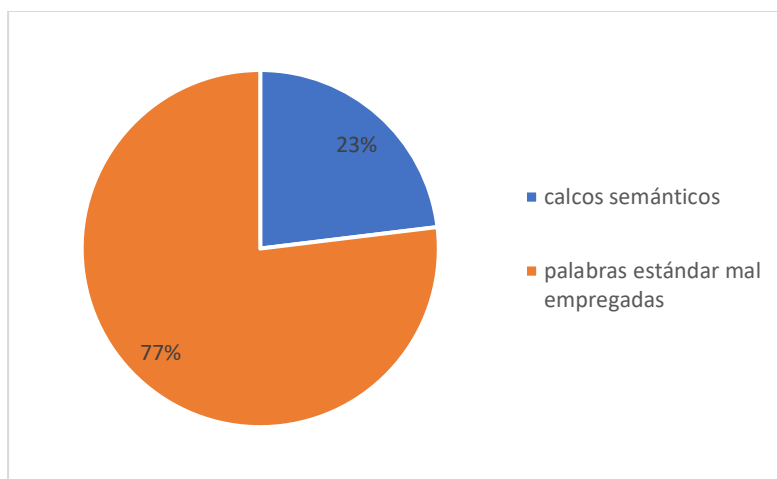


Gráfico 7. Problemas semánticos por lemas

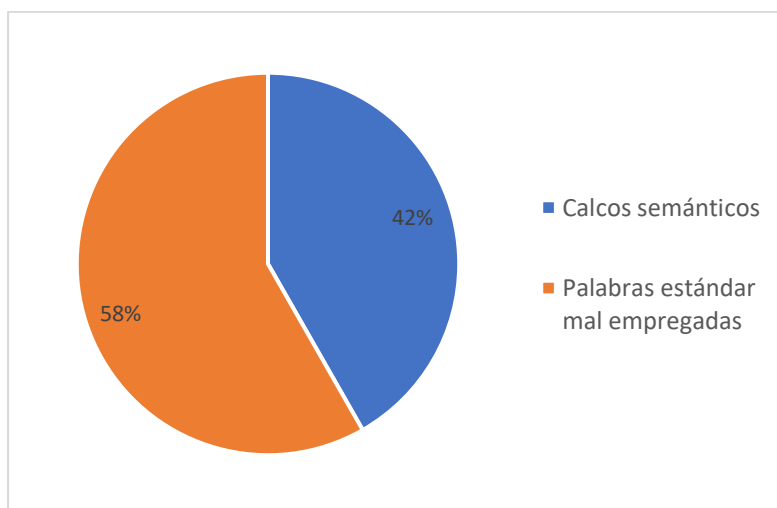


Gráfico 8. Problemas semánticos por formas

En definitiva, o máis destacable en canto á ocorrencia de problemas é, sen dúbida, o gran número de importacións léxicas non adaptadas procedentes do español. Poderíamos pensar que isto non fai máis que constatar a perigosa situación en que se atopa o galego nas xeracións máis novas en que a competencia léxica vai mingando a pasos de xigante. En calquera caso este é un estudo dos 300 textos peor cualificados dos 1000 que contén CORTEGAL e dos moitos máis que se poderían incluír polo que facer conclusións xerais sobre a competencia léxica dos aprendentes é algo arriscado.

Por último, cómpre salientar que a proposta de clasificación establecida non estivo exenta de problemas e limitacións. Unha das dificultades máis evidentes á hora de realizar a clasificación foi decidir que problemas considerabamos e cales quedaban fóra destas categorías. En relación con isto, unha vez decidimos que considerariamos como problemas léxicos as raíces dos verbos irregulares non estándares foi difícil delimitar a lematización destas formas, optando por lematizar precisamente a través da raíz irregular.

Outra dificultade a destacar é o escaso número de materiais ou case nulo, agás poucos traballos sobre CORTEGAL, que existen en galego referidos aos corpus de aprendentes. Tivemos que partir de materiais en inglés, maioritariamente, que non teñen en conta a realidade galega.

Á hora de analizar, tamén foi difícil decidir se analizaríamos os problemas desde unha soa perspectiva ou, tal como fixemos, considerariamos as diferentes posibilidades. Neste senso, seguindo a metodoloxía de CORTEGAL analizamos os problemas atendendo ás diferentes posibilidades. Así, vimos por exemplo que certos dialectalismos se poderían tratar tamén como importacións ou que algúns calcos poderían ser importacións híbridas.

Tamén atopamos dificultades á hora de asignar unha etiqueta no caso dalgunhas formas. Por exemplo, no caso de certas palabras que existen no galego popular, resulta complexo determinar se se xeraron no propio galego ou se proceden do español. Este é o caso de formas verbais (como poderían ser *tuvo* ou *valga*) para as que sería preciso un estudo histórico que non é factible desde o punto de vista práctico, pois a etiquetaxe ten que ser áxil e sinxela.

Está claro que este traballo tería sido máis proveitoso se o tempo permitise analizar os 1000 textos que, na actualidade, compoñen CORTEGAL. Así e todo, non pretendiamos ofrecer unha clasificación definitiva senón presentar unha proposta que se poida tomar en consideración para traballos futuros.

8 Referencias bibliográficas

- Abel, Andrea *et al.* (2016): "An extended version of the KoKo German L1 Learner corpus", en Anna Corazza *et al.* (dirs.), *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016*. Torino: Accademia University Press. <<http://ceur-ws.org/Vol-1749/paper1.pdf>> [Consultado: 29/06/2020]
- Abel, Andrea *et al.* (2014): "KoKo: an L1 Learner Corpus for German", en Nicoletta Calzolari *et al.* (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. Reykjavik: European Languages Resources Association, 2414-2421. <http://www.lrec-conf.org/proceedings/lrec2014/pdf/934_Paper.pdf> [Consultado: 29/06/2020]
- Álvarez, Rosario (coord.): *Tesouro do léxico patrimonial galego e portugués*. Santiago de Compostela: Instituto da Lingua Galega. <<http://ilg.us.es/Tesouro>> [Consultado: 17/07/2020]
- Álvarez de la Granja, María (2018): "Corpus de textos de estudantes galegos (CORTEGAL). Aspectos metodolóxicos", en Marta Díaz *et al.* (eds.), *Actas do XIII Congreso Internacional de Lingüística Xeral*. Vigo: Universidade de Vigo, 55-62. <<http://cilx2018.uvigo.gal/actas/>> [Consultado: 10/06/2020]
- Álvarez de la Granja, María (2020): *Anotación en CORTEGAL*. Documento de traballo interno.
- Cardoso, Adriana *et al.* (2014): "*CUTE: Corpus of Portuguese Undergraduates' Texts. Um recurso para a investigación em escrita académica em português*", en A. Moreno *et al.* (eds.), *Textos Seleccionados, XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto: APL, 169-184.
- Corder, S. P. (1967): "The Significance of Learners' Errors", en *IRAL*, 5, 161-170.
- CORTEGAL = María Álvarez de la Granja/ Ernesto González Seoane (eds.). *Corpus de textos galegos escritos por estudantes no ámbito académico*. Santiago de

Compostela: Instituto da Lingua Galega. <<http://ilg.usc.gal/cortegal/>>
[Consultado: 10/06/2020].

Dagneaux, Estelle / Sharon Denness / Sylviane Granger (1998): "Computer-aided error analysis", *System* 26, 126-174.
<https://www.researchgate.net/publication/222308230_Computer-aided_error_analysis> [Consultado: 9/06/2020]

Díaz-Negrillo, Ana / Jesús Fernández Domínguez (2006): "Error Tagging Systems for Learner Corpora", *Revista española de lingüística aplicada*, 19, 83-102.
<https://www.researchgate.net/publication/28137922_Error_Tagging_Systems_for_Learner_Corpora> [Consultado: 9/06/2020]

Dubert García, Francisco (2004): “*¿Influente ou influínte? Preinscrición e uso da vogal temática verbal da C-III nos adxectivos en -nte na lingua galega*”, en Rosario Álvarez / Henrique Monteagudo (eds.), *Norma lingüística e variación*. Santiago de Compostela: Consello da Cultura Galega / Instituto da Lingua Galega, 223-246. http://consellodacultura.gal/mediateca/pubs.pdf/norma_lingua.pdf
[Consultado: 03/07/2020]

Dubert García, Francisco (2005): “Interferencias del castellano en el gallego popular”, *Bulletin of Hispanic Studies* 83/3, 271-291.
<https://www.researchgate.net/publication/258277513_Interferencias_del_castelano_en_el_gallego_popular> [Consultado: 1/07/2020]

Fernández Salgado, Benigno / Henrique Monteagudo Romero (1995): “Do galego literario ó galego común. O proceso de estandarización na época contemporánea”, en Henrique Monteagudo (ed.), *Estudos de sociolingüística galega. Sobre a norma do galego culto*. Vigo: Galaxia, 99-167.

Freixeiro Mato, Xosé Ramón (2005): “*Non obstante, sen embargo, nembargantes(s) e outros adverbios conectores do galego escrito. Perspectiva diacrónica e uso actual*”, *Verba* 32, 247-274. <<http://hdl.handle.net/10347/3435>> [Consultado: 16/07/2020]

- Granger, Sylviane (2008): "Learner corpus", en Anke Lüdeling / Merja Kytö (eds.), *Corpus linguistics. An international handbook, vol 1*. Berlin, Boston: De Gruyter Mouton, 259-275.
- Martín Peris, Ernesto (dir.) (2008): *Diccionario de términos clave de ELE*. Centro Virtual Cervantes: Instituto Cervantes <https://cvc.cervantes.es/Ensenanza/biblioteca_ele/diccio_ele/diccionario/analisis_serrores.htm> [Consultado: 9/06/2020]
- MERLIN project (2014): *Annotation guidelines*. <<https://merlin-platform.eu/docs/Annotation%20guidelines.pdf>> [Consultado: 29/06/2020]
- Nicholls, Diane (2003): "The Cambridge Learner Corpus—error coding and analysis for lexicography and ELT", en Dawn Archer *et al.* (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, UCREL technical paper number 16, UCREL, Lancaster: Lancaster University, 572-581. <https://www.academia.edu/22736960/The_Cambridge_Learner_Corpus_-_error_coding_and_analysis_for_lexicography_and_ELt> [Consultado: 25/06/2020]
- Parga Valiña, Miruca (2004): "A interferencia lingüística no galego oral", en Rosario Álvarez / Francisco Fernández Rei / Antón Santamarina (eds.), *A lingua galega: historia e actualidade*. Santiago de Compostela: Consello da Cultura Galega / Instituto da Lingua Galega, vol. 1, 547-558. <<http://hdl.handle.net/10347/9892>> [Consultado: 2/07/2020]
- Río, Iria del / Amália Mendes (2018): "Error annotation in the COPLE2 corpus", *Revista da Associação Portuguesa de Linguística*, 4, 225-239. <<https://ojs.apl.pt/index.php/RAPL/article/view/42/44>> [Consultado: 29/06/2020]
- Rodríguez Río, Xusto (2003): "Os préstamos na produción lexicográfica, terminográfica e enciclopédica galega actual", en María Álvarez de la Granja / Ernesto González Seoane (eds.), *A estandarización do léxico*. Santiago de Compostela: Instituto da Lingua Galega / Consello da Cultura Galega, 419-446.

<<http://consellodacultura.org/mediateca/extras/estandarizacion.pdf>>

[Consultado: 1/07/2020]

Romero, Patricia / Manuel C. Núñez Singala (2004): “Interferencias entre galego e castelán na prensa escrita en Galicia”, en Rosario Álvarez / Francisco Fernández Rei / Antón Santamarina (eds.), *A lingua galega: historia e actualidade*, Santiago de Compostela: Consello da Cultura Galega / Instituto da Lingua Galega, vol. 1, 535-545. <<http://consellodacultura.gal/publicacion.php?id=188>>
[Consultado: 2/07/2020]

Sánchez Rufat, Ana (2015): "Análisis contrastivo de interlengua y corpus de aprendientes: precisiones metodológicas", *Pragmalingüística* 23, 191-210. <https://www.researchgate.net/publication/319673058_Analisis_contrastivo_de_interlengua_y_corpus_de_aprendientes_precisiones_metodologicas>
[Consultado: 9/06/2020]

Silva Valdivia, Bieito (1991): “Tipoloxía das manifestacións de contacto lingüístico en Galicia. Algunhas consideracións”, *Cadernos de lingua* 4, 27-38. <<http://publicacionsperiodicas.academia.gal/index.php/Cadernos/article/view/602>> [Consultado: 1/07/2020]

Silva Valdivia, Bieito (2013): “Galego e castelán: entre o contacto e a converxencia”, en Eva Gugenberger / Henrique Monteagudo / Gabriel Rei-Doval (eds.), *Contacto de linguas, hibrididade, cambio: contextos, procesos e consecuencias*. Santiago de Compostela: Consello da Cultura Galega / Instituto da Lingua Galega, 287-314. <<http://consellodacultura.gal/mediateca/documento.php?id=2065>>
[Consultado: 3/07/2020]

Tenfjord, Kari / Paul Meurer / Knut Hofland (2006): “The ASK corpus: A language learner corpus of Norwegian as a second language”, en *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Xénova: European Language Resources Association, 1821–1824. <https://pdfs.semanticscholar.org/af74/17fdebb1d51e08f4b77111d6464fbaa9f6a5.pdf?_ga=2.246502660.1873985775.1585827674-334967235.1585669811>
[Consultado: 29/06/2020]

Wisniewski, Katrin *et al.* (2014): *The MERLIN annotation scheme for the annotation of German, Italian, and Czech learner language*. <<https://merlin-platform.eu/docs/MERLIN-annotation-scheme.pdf>> [Consultado: 29/06/2020]