
Dicionários Monolíngues da Língua Galega*

Iriarte Sanromán, Álvaro (2005) “Dicionários Monolíngues da Língua Galega”, em *Revista Galega de Filoxía*, ISSN 1576-2661. 6 (2005) 51-72; A Corunha; Universidade da Coruña.

Álvaro Iriarte Sanromán
Universidade do Minho

Resumo

O objectivo deste trabalho será mostrar em que medida os principais dicionários monolíngues galegos de que actualmente dispomos no mercado fornecem informação (gramatical, combinatória, pragmática, etc.) suficiente para servirem como ferramentas para ajudar o utilizador a elaborar textos em galego (como língua materna, segunda ou estrangeira). Tentarei quantificar, de maneira aproximada, esta informação, para assim poder justificar respostas a perguntas do tipo: Qual é o melhor dicionário de galego? Que dicionário podemos utilizar para ajudar a aprender galego? Que dicionário nos pode ajudar a escrever em galego? etc.

Palavras-chave

Lexicografia, metalexigrafia, dicionários, dicionários galegos, dicionários de uso, dicionários codificadores, avaliação de dicionários.

Abstract

The aim of this work is to show in what measure the current Galician unilingual dictionaries provide sufficient information (grammatical, combining, or pragmatic information) to serve as a tool to help users to elaborate texts in Galician (as mother tongue, second language or foreign language).

I will attempt to quantify, in an approximate way, this information, for thus be able to justify answers to questions such as: Which is the best Galician dictionary? What dictionary can we use to learn Galician? What dictionary can help us to write in Galician? etc.

Key words

Lexicography, metalexigraphy, dictionaries, Galician dictionaries, usage dictionaries, encoding dictionaries, dictionaries evaluation.

1. Introdução

1.1. Como lexicógrafo, considero que a principal tarefa da lexicografia é registar factos linguísticos e não legislar sobre os mesmos, abandonando as atitudes normativizadoras, puristas ou até xenófobas da lexicografia tradicional. O dicionário está longe de apresentar a língua como uma abstracção. O lexicógrafo compila apenas um repositório de usos consolidados pela norma ou normas culturalmente favorecidas.

Contudo, reconheço a importância do dicionário no processo de construção de uma norma padrão, o papel que o dicionário desempenhou como instrumento de fixação como norma culta de uma determinada variedade linguística. Para além das possíveis

* Um trabalho semelhante, relativo aos dicionários portugueses, intitulado “Dicionários Codificadores”, foi publicado em Sousa, C. e R. Patrício (eds.) (2004), 81-98.

políticas linguísticas (ou ausência das mesmas), o papel do dicionário é simbólico no processo de construção de uma norma padrão. Prova disso é a importância dos dicionários no processo de normalização das línguas minorizadas assim como dos diferentes trabalhos e produtos que hoje se estão a realizar, dentro do que se conhece como “indústrias da língua”, para as principais línguas de cultura (bases de dados lexicais, obras terminográficas científicas e técnicas, bases de dados multilíngues, *thesaurus* para documentação, etc.).

1.2. O objectivo deste trabalho será mostrar em que medida os principais dicionários monolíngues galegos de que dispomos no mercado fornecem informação (gramatical, combinatória, pragmática, etc.)¹ suficiente para servirem como ferramentas para ajudar o utilizador a elaborar² textos em galego (como língua materna, segunda ou estrangeira).

Tentarei quantificar, de maneira aproximada, esta informação, para assim poder justificar respostas a perguntas do tipo: Qual é o melhor dicionário de galego? Que dicionário podemos utilizar para ajudar a aprender galego? Que dicionário me pode ajudar a escrever em galego? etc.

A forma de medir os dados, a própria quantificação dos dados, é questionável. Mas é disso que se trata: É melhor termos um instrumento de comparação (embora imperfeito, aproximado, contestável) do que nenhum. Convidamos desde já outras pessoas a ensaiar outras fórmulas de medida.

Optei por não valorizar de maneira diferente as distintas variáveis (atribuindo-lhes coeficientes de ponderação) para assim não subjectivar ainda mais a análise (principalmente *a posteriori*, com todos os dados já recolhidos). Somente decidimos, *a posteriori*, não contabilizar, no quadro final, os valores relativos aos dados sobre a

¹ No produto final que conforma o dicionário não só se reflectem princípios pertencentes ao nível de análise lexical, pragmático, etc., mas também outros parâmetros para além dos linguísticos, tais como:

- critérios de tipo comercial: o problema do tamanho do dicionário, o tempo de elaboração, o orçamento disponível, etc.;
- critérios de tipo sócio-cultural: as influências ideológicas, eclesiásticas, de orientações linguísticas, de cânones socio-culturais, modas, etc.. O dicionário pode atingir até a categoria de símbolo (pense-se em lugares-comuns como “em todas as casas deveria haver um dicionário, uma, uma gramática, uma bíblia...”);
- critérios de tipo sócio-linguístico: o dicionário pode ser uma ferramenta fundamental nas políticas linguísticas. Um dicionário de português não terá o mesmo valor normativo em Portugal do que em Moçambique, em Timor ou na Galiza, por exemplo, onde, devido às peculiares situações sócio-linguísticas, não se pode pensar em dicionários puramente descritivos entendidos como meros registos do uso real da língua;
- critérios de tipo didáctico (que responda a determinadas necessidades escolares ou de utilizadores de línguas maternas diferentes);
- critérios estéticos; etc.

² A característica mais importante dos dicionários codificadores é que deverão fornecer ao utilizador mais informação morfo-sintáctica, semântica e pragmática do que um dicionário descodificador, uma vez que na actividade descodificadora aplicamos estratégias de tipo textual ou pragmático que nos permitem inferir o significado de determinada palavra ou combinação lexical, estratégias das quais não dispomos no momento da codificação linguística

pronúncia, uma vez que os poucos valores recolhidos referem-se apenas à informação sobre a existência de dois graus de abertura nas entradas correspondentes ao **e** e ao **o**.

1.3. Para realizarmos a análise dos dicionários, seleccionámos os seguintes lemas, tomados duma listagem dos 1000 lemas mais frequentes do *Corpus de Referencia do Galego Actual* (GORGA)³.

a) 15 palavras lexicais ou plenas⁴, das quais:

- Os 7 primeiros substantivos⁵: *ano, vez, día, vida, home, parte, casa*.
- 7 verbos⁶: *poder, ir, facer, dicir, saber, querer, deber*.
- 4 adjectivos⁷: *primeiro, novo, mellor, maior*.
- 2 advérbios: *cando, ben*⁸.

b) 5 palavras gramaticais ou funcionais⁹:

³ Quero agradecer ao Pedro Dono (na altura, do Centro de Estudos Galegos da Universidade do Minho) o envio da listagem dos 1000 lemas mais frequentes do *Corpus de Referencia do Galego Actual* (CORGA). Sobre o CORGA, vd. <http://corpus.cirp.es/corga.info.html>

⁴ Não queremos entrar aqui no problema das "partes da oração". Limitar-nos-emos a distinguir dois grandes grupos: o das palavras lexicais ou plenas e o das palavras gramaticais ou funcionais, entendendo estas últimas como as palavras que cumprem, em parte ou inteiramente, funções meramente estruturais ou gramaticais e que não têm um significado lexical ou que é difícil de precisar (vd. Lewandowski, 1986: s.v. **palabras funcionales**, e Xavier e Mateus (org.) 1991: s.v. **palavra funcional**).

⁵ Para a contagem das frequências dos substantivos considerámos as formas no singular e no plural (ano + anos; etc.)

⁶ Excluimos *ser, estar, ter* ou *haber*, também com altos índices de frequência, porque consideramos que pertencem ao grupo das palavras funcionais (vd. *infra* nota 9). Para a contagem das frequências dos verbos contabilizámos todas as formas da palavra (Lyons 1980: 20) registadas entre as 1000 formas mais frequentes do CORGA e não apenas a forma de citação infinitivo (*facer, fago, fixen, etc.*): *poder* 43037 ocorrências; *ir* 42381 ocorrências (mesmo sem contabilizar formas como *foron, fose, etc.*, que poderão ser também do verbo *ser*, o verbo *ir* está entre os 6 mais frequentes com 23142 ocorrências); *facer* 41590 ocorrências; *dicir* 34872 ocorrências; *saber* 25262 ocorrências; *querer* 15727 ocorrências; *deber* 9571 ocorrências.

⁷ Para a contagem das frequências dos adjectivos tivemos em conta as formas no singular, plural, no masculino e no feminino (*primeiro, -a, -os, -as*; etc.). Excluimos *galego*, com o maior índice de frequência, devido à sua frequente função de adjectivo. Outros adjectivos, com maiores índices de frequência, como *mesmo, mellor, só, etc.* foram excluídos por terem muito frequentemente outras funções para além da de adjectivo.

⁸ Em termos de frequência, *cando, ben* e *aínda* não são dos primeiros advérbios da listagem. Estão aqui incluídos porque não podemos considerá-los como sendo apenas palavras funcionais como outros advérbios num lugar mais alto na lista (*non, máis, xa, etc.*).

⁹ Das numerosas palavras gramaticais que encabeçam a lista dos 1000 lemas mais frequentes do *corpus*, escolhi as primeiras, excluindo a forma *a* (que poderia ser artigo, pronome, substantivo masculino e preposição e que, lexicograficamente, na maior parte destes casos já está representado pelo lema *o*) assim como *do* (*de+o*) e *da* (*de+a*), que, de alguma maneira, já estão representados, embora separadamente, por *de* e por *o*. Embora a forma *o* também possa ter as funções de substantivo masculino, artigo definido e pronome, a sua frequência é mais do que duas vezes superior à do lema *non*, que seria o seguinte a ser incorporado, para além de representar lexicograficamente as formas *o, a, os, as*). Incluímos também o lema *ser*, que, como *forma de citação* (Lyons, 1980: 20) das várias formas de palavra (*son, é, era, foi, etc.*) registadas entre as 1000 formas mais frequentes, estaria situada em 6º lugar em termos de frequência no CORGA:

Orde	Forma	Frecuencia	Porcentaxe	Porcentaxe acumulado
------	-------	------------	------------	----------------------

- *de, que, e, o, ser.*

Após a selecção dos 25 lemas, foram contabilizados nas correspondentes entradas a totalidade das acepções registadas, independentemente da categoria gramatical (assim, na entrada **deber**, por exemplo, foram contabilizadas as acepções correspondentes à categoria verbo, mas também as da categoria substantivo) e também independentemente do tratamento polissémico ou homonímico que cada dicionário possa dar aos termos (**poder**, *v.tr., i. e s.m.* vs. **poder**¹, *v.tr. e i.*; **poder**², *s.m.*; etc.)¹⁰.

Não serão contabilizadas formas com qualquer tipo de divergência gráfica, como maiúsculas (**ir** vs. **Ir**), prefixos (**e** vs. **e-**), acentos gráficos (**o** vs. **ó**), etc. Mas sim serão contabilizadas as formas femininas (**primeiro, primeira**) ou plurais (**deber, deberes**) registadas.

1.4. Foram objecto de análise dois tipos de dicionários de língua, os dicionários de grande formato, ou “gerais”, e os dicionários “manuais” (todos disponíveis no mercado)¹¹:

1.4.1. Dicionários Gerais¹²:

ALONSO ESTRAVÍS, I. (1995) *Dicionário da Língua Galega*. Santiago: Sotelo Blanco [aqui: *Sotelo Blanco*].

CARBALLEIRA ANLLO, X. M. (coord.) (2000) *Gran Dicionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Gran Xerais*].

LEDO CABIDO, B. (dir.) (2004) *Dicionario de Galego*. Vigo: Ir Indo Edicións [aqui: *Ir Indo*].

PENA X.A. (dir.) (2004) *Gran Dicionario Cumio da Lingua Galega*. Vigo: Edicions do Cumio [aqui: *Gran Cumio*].

1.4.2. Dicionários manuais:

GARCÍA, C. e M. GONZÁLEZ GONZÁLEZ (dir.) (1997) *Diccionario da Real Academia Galega*. Vigo: RAG [aqui: *RAG*].

1	de	733000	4.154	4.154
2	que	667031	3.780	7.934
3	a	627791	3.558	11.49
4	e	529841	3.002	14.49
5	o	485749	2.753	17.25
5b	ser	± 287264		

¹⁰ Independentemente dos vários critérios de diferenciação que se possam usar (etimológicos, semânticos, morfo-sintácticos), a distinção entre polissemia e homonímia nem sempre é fácil de fazer, nomeadamente na prática lexicográfica. De qualquer maneira, isto não justifica a disparidade de critérios que encontramos, como por exemplo, no *Ir Indo*: **deber** (*v. e s.*); **dicir** (*v. e s.*); vs. **poder**¹ (*v.*) e **poder**² (*s.*); **saber**¹ (*v.*) e **saber**² (*s.*); **ser**¹ (*v.*) e **ser**² (*s.*), etc.

¹¹ Seguimos, para esta categorização, González Seoane (2003: 179-183), para quem, no âmbito da lexicografia galega seriam “dicionários gerais” as obras com mais de 70.000 artigos e “dicionários manuais”, as obras entre 25.000 e 50.000 entradas.

¹² Quero agradecer ao Carlos Pazos (do Centro de Estudos Galegos da Universidade do Minho) o fornecimento da maior parte dos dicionários galegos aqui analisados.

- PENA X.A. (dir.) (2004) *Diccionario Cumio da Lingua Galega*. Vigo: Edicions do Cumio [aqui: *Cumio*].
- CARBALLEIRA ANLLO, X. M. (coord.) (2004) *Diccionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Xerais*].

2. Análise da microestrutura dos dicionários

Embora o trabalho de marketing das editoras continue a utilizar como chamariz o número de vozes recolhidas (e, como lexicógrafo, posso dizer que sei bem como isso é importante para as casas editorais), um dicionário não deverá apenas ser avaliado pelo número de entradas que recolhe – a sua **macroestrutura** – mas pelo tratamento que se dá – na **microestrutura** – às vozes recolhidas.

Como já indiquei noutra altura (Iriarte Sanromán 2001: 299), conhecer uma unidade lexical de uma língua implica principalmente:

- Conhecer a sua forma oral e escrita (pronúncia e ortografia).
- Conhecer as suas variações formais (género, número, etc.) assim como a capacidade para formar novas unidades (derivação, composição, etc.).
- Conhecer as suas capacidades combinatórias, as relações possíveis com outras unidades no co-texto (combinações livres, colocações, frasesmas, etc.).
- Conhecer os seus significados ou acepções, assim como as possibilidades de uso pragmático-contextual e retórico (contextos de uso, registos, domínios, usos metafóricos, fórmulas de rotina, variantes diacrónicas, geográficas, estilísticas, etc.).

Em princípio, quanto mais informação deste tipo estiver consignada e convenientemente etiquetada¹³ num dicionário, melhor será a obra lexicográfica.

Eis alguns elementos da microestrutura que seguidamente analisaremos nos nossos dicionários:

- (1) Definições e acepções do lema
- (2) Informação sintagmática (número de subentradas em forma de expressões pluriverbais formadas pelo lema mais outra ou outras palavras)
- (3) Transcrição(-ções) fonética(s) ou figurada(s) do lema.
- (4) Exemplos e abonações.
- (5) Etiquetagem gramatical (informações, restrições ou explicações ortográficas, morfológicas, sintáticas, semânticas, lexicais, etc.).

¹³ Apesar das críticas de que, por vezes, são alvo as etiquetas utilizadas nos dicionários, pensamos que a definição mais útil que se pode dar de uma determinada unidade lexicográfica, especialmente nos chamados dicionários de produção ou codificadores (unilingues ou bilingues), será a informação fornecida por um sistema de etiquetagem o mais completo possível, que ultrapasse sem nenhum tipo de receio o imanentismo gramatical a que nos habituou a linguística do século XX.

Neste sentido, não pensamos que um sistema de etiquetagem seja, *a priori*, mais imperfeito ou incompleto do que uma definição. Para além de podermos conceber a definição como fazendo parte do conjunto de etiquetas, podem existir bons e maus sistemas de etiquetagem e boas (em lexicografia, no sentido mais pragmático de 'úteis' do que no sentido de 'científicas') e más definições.

- (6) Etiquetagem enciclopédico-cognitiva (áreas de conhecimento, etc.).
- (7) Etiquetagem pragmático-retórica (restrições e informações pragmáticas, retóricas e/ou contextuais).

É importante assinalar que estes diferentes tipos de informação não terão necessariamente de ser recolhidos e etiquetados em compartimentos estanques. Muita desta informação estará disseminada na microestrutura e não propriamente agrupada em informação sintáctica, semântica, enciclopédica ou pragmática, tal como a acabamos de apresentar. Assim, por exemplo, muitas vezes será difícil estabelecer os limites entre o que é uma definição e as etiquetas que delimitam o uso co-textual (sintáctico ou combinatório) e contextual (pragmático ou enciclopédico) de uma determinada unidade lexicográfica. Outro exemplo disto são etiquetas como FIG. ou POR EXTENSÃO: são etiquetas com informação gramatical (no sentido lato em que aqui entendemos gramatical) ou com informação pragmática-retórico como COLOQ., VULG. ou DESUS.?

2.1. Número de acepções

Nem sempre é fácil delimitar e contabilizar o número de acepções. Esta dificuldade reflecte a natureza não discreta do fenómeno do significado, uma vez que estes, os significados, não são como as coisas, entidades separáveis, contáveis (*cf.* Silva 1997: 587-588)¹⁴.

Por outro lado, mais importante do que o número de acepções será o tratamento e o desenvolvimento que é dada a cada uma delas: uma série de 10 ou 20 sinónimos separados por ponto e vírgula não terão o mesmo valor que 10 ou 20 acepções delimitadas numericamente, com definições em forma de paráfrases bem desenvolvidas, informações gramaticais, restrições de uso, etc.¹⁵

Eis a seguir o gráfico correspondente ao número de acepções contabilizadas na totalidade dos lemas seleccionados:

¹⁴ Relacionado com isto está a questão, que aqui deixaremos de lado, do tratamento que os dicionários dão aos diferentes lemas, como palavras polissémicas ou homónimas: *costa* e *costas* seriam contabilizados aqui independentemente de fazerem parte ou não da mesma entrada lexicográfica.

¹⁵ No excelente (e belo) *Dicionário Houaiss da Língua Portuguesa*, por exemplo, encontramos acepções e subacepções relacionadas mediante números, numa estrutura semelhante à de um índice (1, 2, **2.1.**, **2.2.**, **2.3.**, **2.3.1.**, 2.4., etc.) (*cf.* Houaiss (coord.) (2001), *passim*).

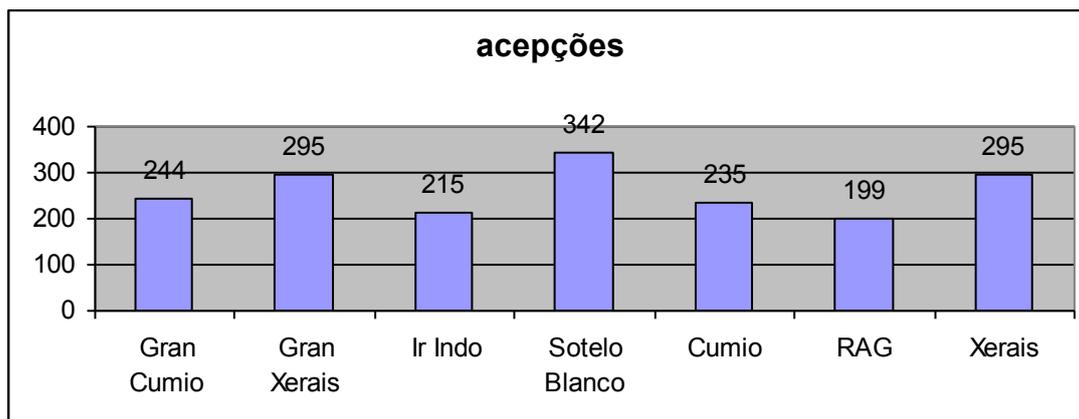


Gráfico 1
Acepções

2.2. Expressões pluriverbais (informação sintagmática ou combinatória)

Qualquer palavra, signo, letra (ou conjunto de palavras, signos ou letras) poderia constituir uma entrada num dicionário. Contudo, são claras as vantagens na utilização como lema da unidade “palavra”¹⁶, nomeadamente nos dicionários tradicionais em formato não electrónico¹⁷, onde qualquer tipo de unidade pluriverbal deverá ser registada, em forma de subentrada, sob uma ou várias entradas¹⁸ das várias palavras lexicais que compõe a expressão pluriverbal, correspondendo assim ao que Cowie (1983: 99) chama “expectativas conservadoras dos usuários comuns dos dicionários”.

Tão importante como o número de entradas e o número de acepções registadas é a quantidade de subentradas relativas a expressões pluriverbais formadas pelo lema combinado com outras palavras. A informação sobre combinatória lexical é uma mais-valia importante em qualquer dicionário.

Por outro lado, queremos chamar a atenção para o facto de que, muito frequentemente, os dicionários apresentam como acepção de um lema o que em rigor é o significado desse lema combinado com outras palavras (“ter bom ouvido”, como acepção de **ouvido**, por exemplo) (Calderón 1994: 58). Tais acepções são, de facto, combinações lexicais que deveriam ser registadas em forma de subentradas.

Assim, serão contabilizadas aqui “pseudo-acepções” como

¹⁶ No sentido em que um falante corrente entende intuitivamente o termo: conjunto delimitado por dois espaços em branco, espaço e sinal de pontuação ou espaço e hífen. Werner (1982: 224-229) apresenta vários argumentos para defender que o lema deverá corresponder à unidade palavra.

¹⁷ Nos dicionários em formato electrónico podem utilizar-se vários processos de pesquisa e de recuperação de informação (por exemplo, por meio dos operadores lógicos *e* e *ou*) para procurar unidades pluriverbais.

¹⁸ O tratamento não será o mesmo para os *frasesmas completos*, para as *colocações*, etc. (vd. Iriarte Sanromán 2001: § 4.5 e § 5.4).

“... 10. Cos adverbios *ben* ou *mal*, cadrar ou non unha cousa con outra; casar, harmonizar” (Gran Xerais, s.v. **dicir**),

pois, em rigor, não estamos perante uma acepção da palavra *dicir*, mas perante o significado da expressão *dicir ben* ou *dicir mal*.

Apresentamos a seguir um gráfico que, de alguma maneira, tenta quantificar o tratamento que cada dicionário dá à importante questão da combinatória lexical, indicando explicitamente que o valor (acepção) da palavra vem dado pela combinação da mesma com outras palavras. Se for o caso, serão contabilizadas as diferentes acepções das unidades pluriverbais¹⁹. Como podemos apreciar no gráfico, o *Gran Xerais* destaca-se no tratamento dado à parte sintagmática:

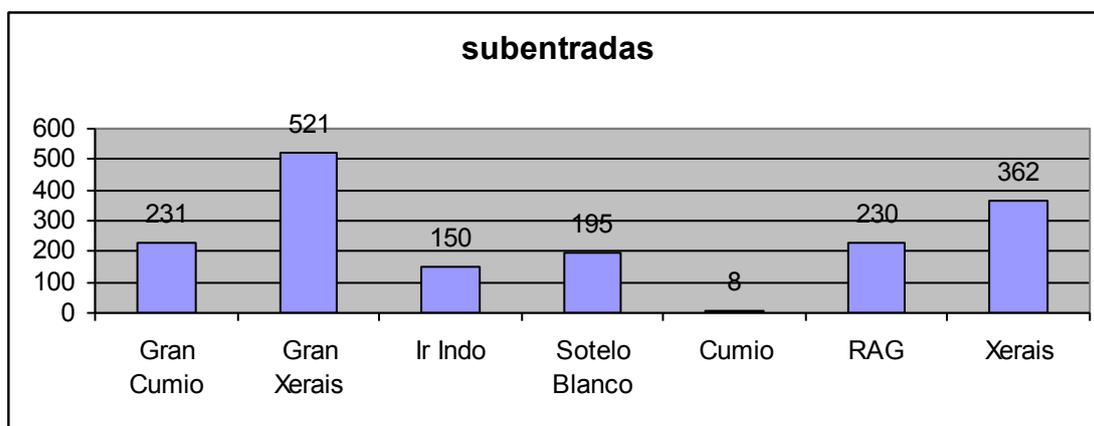


Gráfico 2
Subentradas

2.3. Transcrição(-ções) fonética(s) ou pronúncias figurada(s)

A transcrição fonética dos grandes dicionários de língua deveriam registar as principais pronúncias existentes, sem grandes pruridos normativizadores ou de ortoépia:

tellado [te'LaDo / te'yaDo] s.m. 1. ...

Poderão também registar-se variantes e até formas claramente afastadas daquilo que poderíamos considerar galego padrão. A(s) forma(s) considerada(s) padrão ficariam marcadas pelo facto de não ter nenhum tipo de etiqueta:

¹⁹ Não contabilizaremos as formas pluriverbais com direito a entrada diferenciada em alguns dicionários, como por exemplo *home-bon* (vs. *home bon*) ou *ben-falado* (vs. *benfalado*), etc. porque iria desvirtuar, em termos quantitativos, o resultado final, uma vez que, no caso das entradas independentes suporia contabilizar novamente a informação gramatical correspondente (sobre a categoria do lema), etc., informação que não seria contabilizada se a construção estiver registada como subentrada.

De facto, a prática de grafar com hífen estas formas compostas não passa de uma convenção puramente gráfica que parece não visar senão uma solução para o problema da lematização destes compostos. A questão das unidades pluriverbais pouco tem a ver com a tradição ortográfica. Sobre o assunto, *vd.* Mathieu-Colas (1994), Catach (1981), Herculano de Carvalho (1979: 506-507, nota 9).

gato ['gato / 'gato (DIALECTAL, POP.)] ...
dez ['deT / 'des (GALEGO OCCIDENTAL)].

Evidentemente, nos dicionários em formato electrónico e com recursos multimédia, esta transcrição fonética poderá ser, melhor do que substituída, acompanhada do registo em áudio da pronúncia das palavras dicionarizadas.

Neste ponto surpreende a falta generalizada de informação sobre a pronúncia Sabia que não existia nos dicionários galegos registo sistemático da pronúncia, seja em forma de transcrição fonética, seja em forma de pronúncia figurada. Contudo, esperava, quando menos, informação sobre o grau de abertura da vogal tónica, sobre fenómenos de metafonia com alternância vocálica em formas como **novo** vs. **nova**, etc.

Em termos quantitativos, como já advertimos na introdução, não contabilizaremos esta variável no quadro final, uma vez que os poucos valores recolhidos referem-se apenas à informação sobre a existência de dois graus de abertura nas entradas correspondentes ao **e** e ao **o**. Mas, não queríamos deixar de apresentar este ponto, e os valores recolhidos no gráfico, justamente por se destacarem negativamente em todos os dicionários:

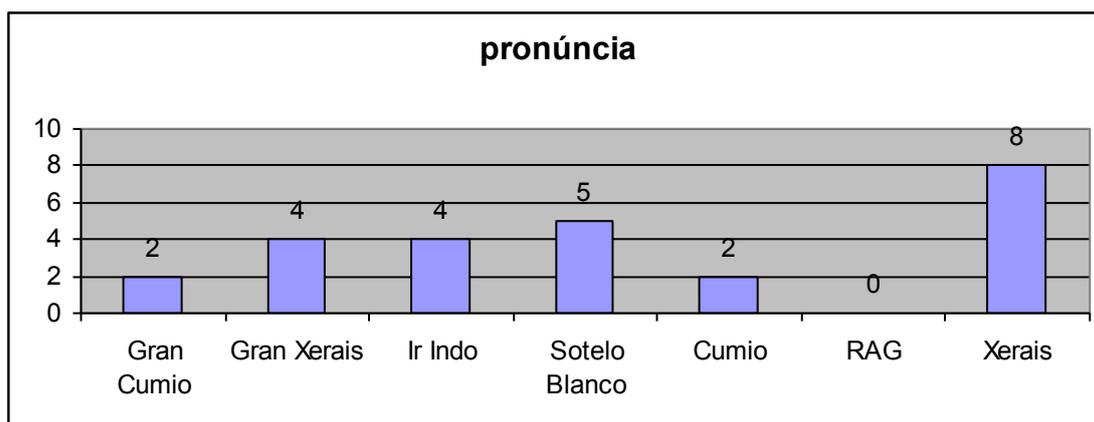


Gráfico 3
Pronúncia

2.4. Exemplos e abonações

Não temos grandes dúvidas de que os exemplos e as abonações podem ser muito ricos em informação morfológica, sintáctica, combinatória, semântica, enciclopédica, pragmática, estilística, etc. Os exemplos são cruciais no uso do dicionário como ferramenta para a codificação. Contudo, devem ser usados com reservas pelo lexicógrafo e só como complemento ao sistema de etiquetagem, dado que podem transformar-se perigosamente numa espécie de “*cajón de sastré*” para onde vai parar tudo aquilo que não sabemos como tratar lexicograficamente.

No plano puramente metalexigráfico, assim como na prática lexicográfica mais recente, parece haver uma clara tendência para defender o uso dos exemplos

procedentes, com a ajuda de ferramentas informáticas, de *corpora* variados, mais do que procedentes da intuição linguística ou subjectividade do lexicógrafo. Isto não justifica, porém, uma certa onda de desconfiança sobre a competência linguística do lexicógrafo (e do linguista em geral) que se pode encontrar em alguns dos autores para os quais a maior objectividade dos dados tomados dum *corpus* (em definitivo de outros informantes, anónimos ou não) significa também uma maior garantia face aos dados subjectivos, fantasiosos, inventados, limitados ou redutores do “informante” lexicógrafo²⁰.

Há também autores que, embora não neguem a utilidade de um *corpus* para confirmar a existência real e os usos pragmáticos dos exemplos empregues no dicionário, assim como os usos gramaticais e semânticos das acepções recolhidas, defendem as vantagens do exemplo construído *ad hoc*, principalmente nos chamados dicionários de produção. Como no caso da contagem das acepções e subacepções, nem sempre é fácil delimitar e contabilizar o número de exemplos: como é que devemos fazer a contagem de casos como: *fazer um edifício, fazer uma piscina, fazer um estádio*, por um lado, e *fazer um edifício, uma piscina, um estádio*, por outro?

Eis a seguir o gráfico correspondente ao número de exemplos registados na totalidade dos lemas estudados. Como vemos, é o dicionário da *RAG* que se destaca quanto ao número de exemplos recolhidos:

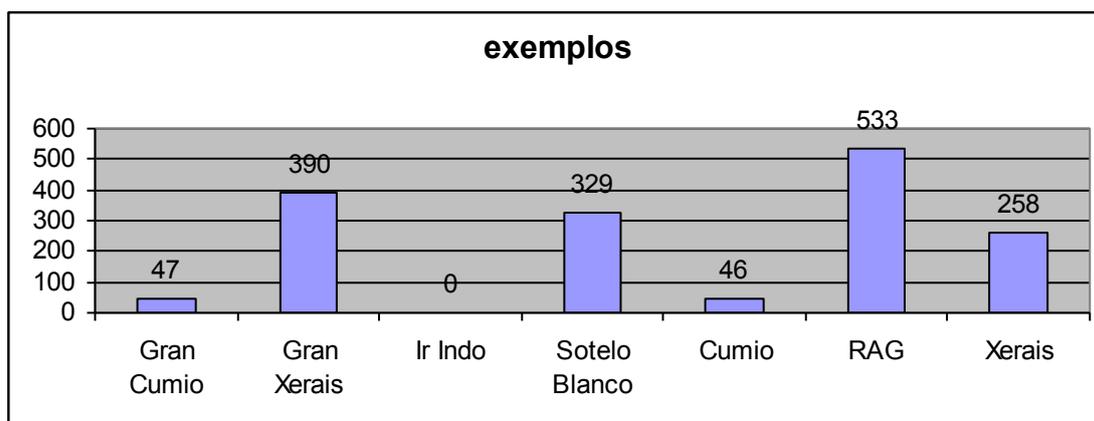


Gráfico 4
Exemplos

2.5. Informação gramatical

²⁰ A polémica é antiga e não só aplicada à elaboração de dicionários, mas à descrição linguística em geral:

La querelle est ancienne: Vaugelas s'en tenait aux usages observés à la Cour et dans certains milieux parisiens, Arnauld et Lancelot (auteurs de la grammaire dite de Port-Royal) illustraient leurs analyses d'exemples inventés par eux-mêmes; les structuralistes, surtout du courant distributionnel, partent de corpus, les générativistes préfèrent le plus souvent s'appuyer sur leur propre sentiment linguistique de la langue dont ils cherchent à “découvrir” la grammaire. (Besse & Porquier 1984: 13-14).

Analísamos aqui a presença de qualquer tipo de informação gramatical nos dicionários, nomeadamente informações ou restrições de carácter ortográfico, morfológico, sintáctico, semântico ou lexical relativa a cada lexema ou combinação de lexemas. Contabilizámos neste grupo:

- A informação sobre a categoria (ou subcategoria) gramatical do lema²¹;
- Informação ortográfica sobre as unidades lexicográficas;
- Informação morfológica, como, por exemplo a formação do plural ou do feminino, modelos de conjugação verbal, etc.;
- Informação sobre a regência verbal (*falar em/falar de*, etc.);
- etc.²²

A seguir apresentamos um gráfico onde tentamos quantificar a informação gramatical recolhida na totalidade dos lemas estudados.

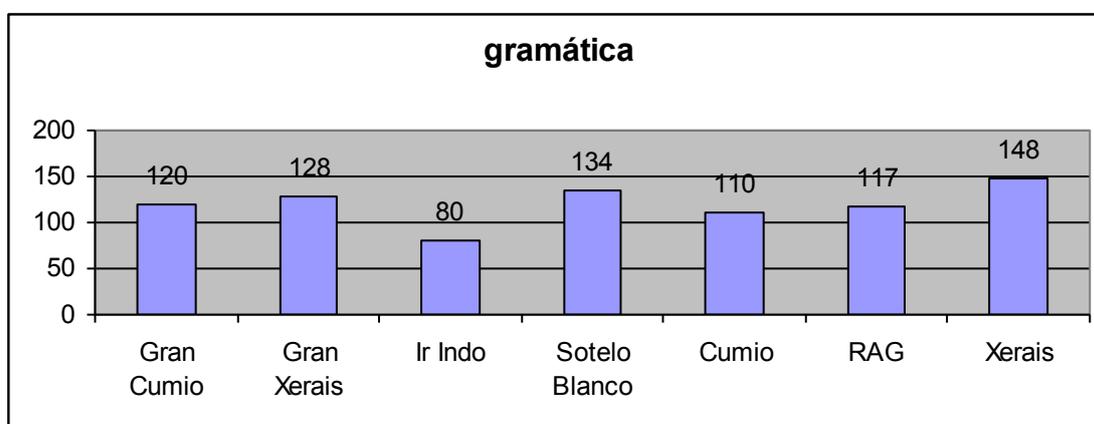


Gráfico 5
Informação gramatical

2.6. Informação enciclopédico-cognitiva

Analísamos aqui a informação relativa ao campo ou sistema conceptual, área de conhecimento e ao marco de referência.

²¹ A indicação da categoria gramatical do lema é uma presença nas obras lexicográficas a que os dicionários existentes nos têm habituados e que a maior parte dos autores considera imprescindível. Noutro lugar já falámos (Iriarte Sanromán 2003) sobre a pouca utilidade que as categorizações gramaticais podem vir a ter para os utilizadores, em grande parte desconhecedores deste tipo de terminologia (para além de que muito frequentemente os sistemas de etiquetagem actualmente existentes nos dicionários são excessivamente redutores para os próprios linguistas: um verbo transitivo pode intransitivizar-se, por exemplo, ou um verbo que rege preposição pode não ser considerado como sendo transitivo, etc.).

Por razões semelhantes, não foram contabilizadas aqui etiquetas como *loc.*, *loc. adv.*, etc.

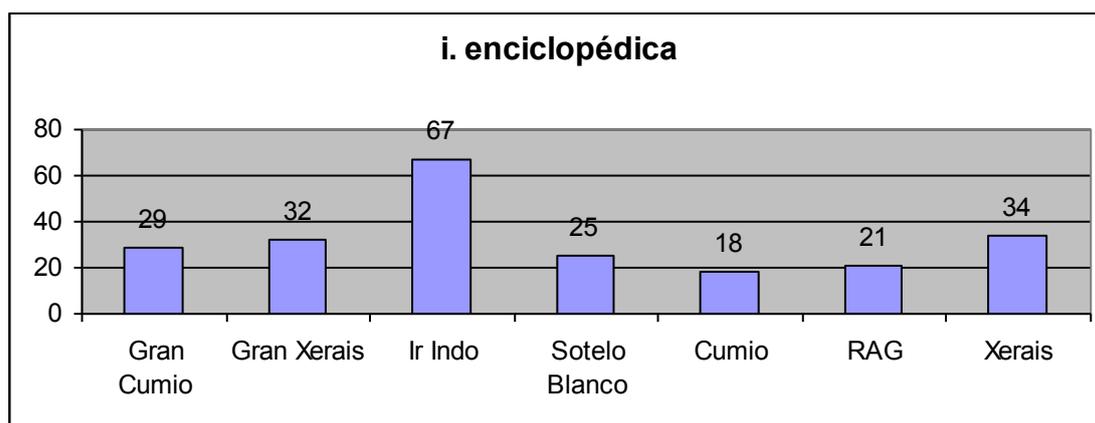
²² Decidimos não contabilizar sinónimos e antónimos porque alguns dicionários fornecem este tipo de informação claramente destacadas, com etiquetas como SIN. ou ANT., mas outros dicionários incorporam-nos nas próprias definições.

Nos dicionários actuais, este tipo de informação parte muitas vezes da intuição do próprio lexicógrafo mais do que de uma classificação produto de uma sistematização rigorosa²³. Contudo, esta é uma informação extremamente importante uma vez que a mudança da área de conhecimento não é irrelevante na descrição lexicográfica de uma palavra porque vai implicar sempre uma mudança no tipo de definição utilizada. Não será a mesma definição a que se utilizará para o sentido de *ovo* como ‘alimento’, dentro do marco de referência da linguagem quotidiana ou da culinária, que a definição de *ovo* para o sentido de ‘célula’, dentro do marco de referência da Biologia.

Por outro lado, queremos insistir novamente no facto de que a informação que se fornece com estas etiquetas não se encontra em compartimentos estanques e perfeitamente delimitados. Assim, este tipo de informação, nem sempre é fácil de distinguir da própria definição (linguística ou enciclopédica?) ou do que aqui chamaremos “etiquetagem de tipo pragmático-contextual”. Por exemplo, a informação que fornecem as etiquetas sobre usos tecnolectais ou dialectais, ou até sobre o uso familiar, popular ou figurado pode ser de tipo pragmático-contextual ou enciclopédico-cognitivo.

No gráfico que a seguir apresentamos, contabilizámos as etiquetas relativas às áreas de conhecimento, matéria ou especialidade, usos tecnolectais ou dialectais, com eventuais etiquetas como: OURENSE, BIOLOGIA, BÍBLIA, CULINÁRIA, ESCOLAR, etc.

Não considerámos aqui a abundante informação enciclopédica eventualmente contida na própria definição (uma vez que seria bastante difícil de quantificar, e até de separar da própria definição linguística²⁴), salvo quando equivalha claramente a uma etiqueta, como por exemplo: “*Na Galiza, ...*”; “*En filosofía, ...*”; “*..., segundo o rito católico, ...*”; “*En determinados xogos,...*”; etc. (frequentemente delimitados, aliás, entre parênteses ou vírgulas).



²³ Neste sentido, escreve González (1996: s.v. **reseñas**): “si consultamos el DRAE [Diccionario de la Real Academia Española] por ramas del saber, encontraremos 844 entradas en el campo milicia y sólo 3 en el campo ingeniería [...]”.

²⁴ Definições claramente de tipo funcional (para que serve ou quando se utiliza aquilo que se está a descrever), definições contendo nomes próprios, etc. são exemplos de conteúdos claramente enciclopédicos. Sobre a distinção entre obras de carácter lexicográfico e obras de carácter enciclopédico, *vd.* Haiman (1980) e Frawley (1981).

Gráfico 6 Informação enciclopédico-cognitiva

2.7. Informação pragmático-contextual e retórica

Contabilizámos aqui informações como:

- actos de fala: a função ou intenção retórico-comunicativa (pense-se por exemplo nas etiquetas do tipo “*usado para*”, “*emprégase como fórmula de cortesia*”, etc.). É fundamental registar no dicionário de língua os usos de uma determinada construção, por exemplo: usado para cumprimentar, despedir-se, oferecer, convidar, pedir, rejeitar, agradecer; para pedir ou dar informações, para iniciar uma conversa ou estabelecer turnos de palavra; para pedir conselho, dar uma ordem, chamar a atenção, pedir autorização para, negar-se a, ameaçar; para exprimir sentimentos, gostos, opiniões, atitudes, etc.

- Informação sobre variações formais deliberadas (estruturais ou textuais), etiquetada na tradição lexicográfica com etiquetas como: COLOQ., POÉTICO, FORMAL, etc.

- Informação relativa às variações sócio-linguísticas e etno-linguísticas, etiquetadas na tradição lexicográfica normalmente com abreviaturas como: CAL., FAM., INF., POP.

- Variações diacrónicas, registadas nos dicionários com etiquetas como ARC., ANTIGO, NEOLOG.

- etc.

Como acontece com a informação de tipo enciclopédico-cognitivo, este tipo de informação parte muitas vezes da intuição do próprio lexicógrafo, visto que é muito difícil delimitar, por exemplo, o que é um uso familiar, popular ou figurado²⁵. Há, todavia, tentativas de classificação sistematizada dos diferentes tipos de etiquetas lexicográficas correspondentes aos diferentes contextos e variantes de uso, como a de Hartmann (1983), que estabelece variantes correspondentes a nove contextos de uso diferentes. O próprio autor reconhece que a distinção entre um tipo e outro muitas vezes é difícil de estabelecer.

Neste último gráfico antes das conclusões, contabilizámos este tipo de informação pragmática nos dicionários analisados:

²⁵ Muitas vezes, os adjectivos ou verbos que, juntamente com um substantivo, conformam uma colocação são apresentados lexicograficamente como sendo acepções ou sentidos figurados (ou impróprios, ou translaticios). Assim, geralmente são consideradas como variações ou manifestações de um único sentido figurado ou metafórico acepções cujo valor só é actualizado quando combinado com outras palavras. É o caso de numerosas colocações (*vd.* Alonso Ramos 1993: 169).

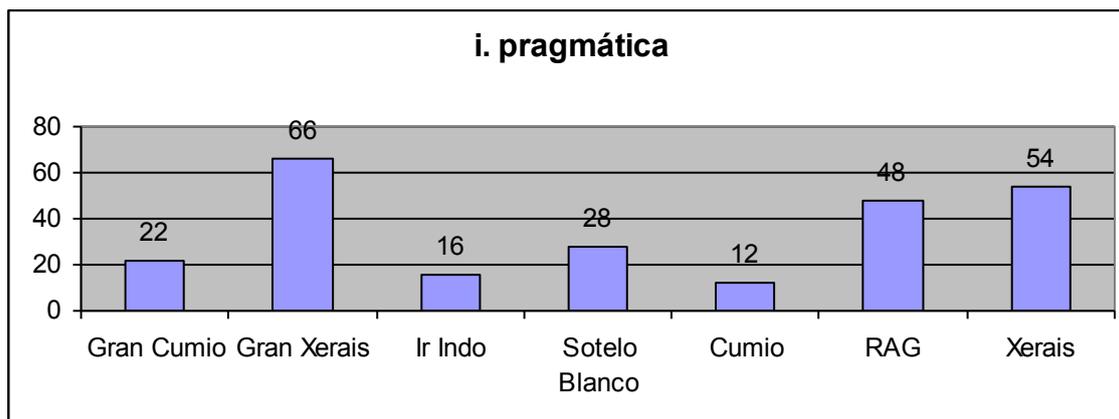


Gráfico 7
Informação pragmático-contextual e retórica

5. Conclusões.

5.1. Os melhores dicionários

Apresentaremos a seguir dois últimos gráficos com a soma dos valores de todos os quantitativos apresentados acima. Para chegar aos números apresentados, decidimos igualar a 100 o maior valor registado (e não o somatório dos valores registados na totalidade dos dicionários analisados) e a partir dele estabelecer as percentagens para os outros dicionários²⁶.

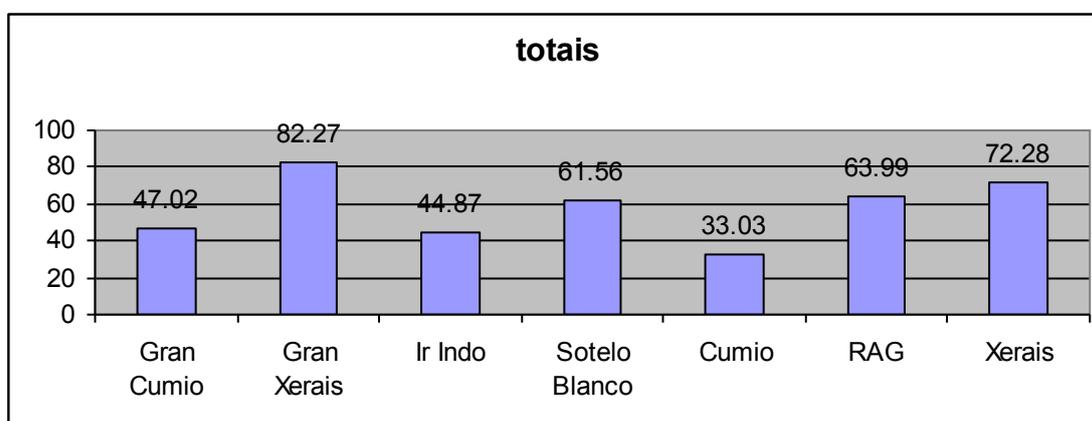


Gráfico 8
Totais

Reordenamos agora os resultados para melhor se poder apreciar e comparar os resultados finais, e descobrir quais são “os melhores dicionários”:

²⁶ Quero agradecer ao Engenheiro José João Almeida (do Departamento de Informática da Universidade do Minho) as suas sugestões e comentários para a elaboração do quadro com os valores totais final. Evidentemente, os eventuais erros ou faltas de exactidão são da minha inteira responsabilidade.

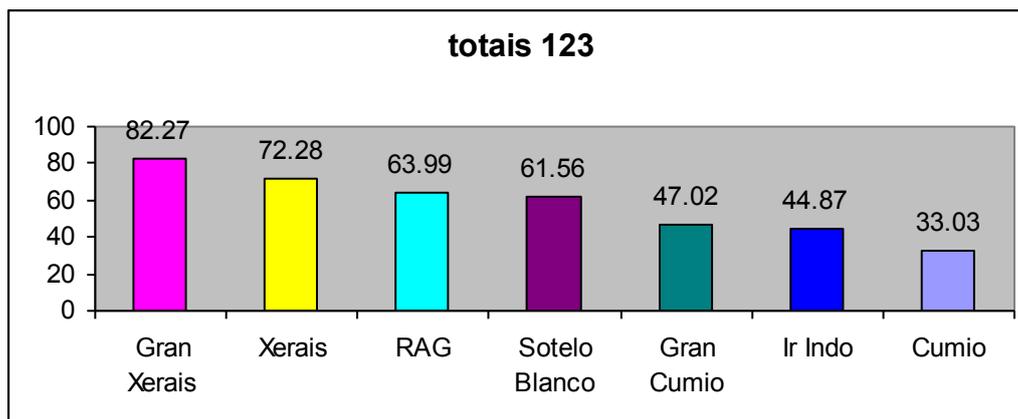
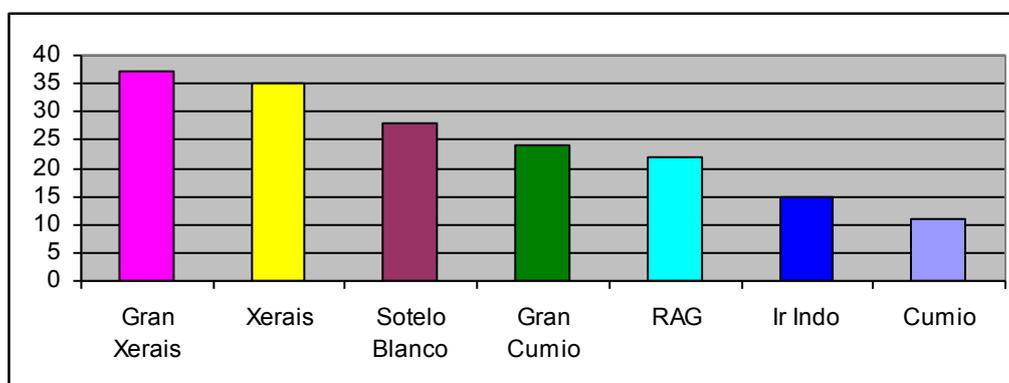


Gráfico 9
“O melhor dicionário é...”

Evidentemente, e como já advertíamos no início deste trabalho, este resultado final será sempre questionável (a própria quantificação dos dados já é questionável). Contudo, pensamos que o resultado pode ser uma boa ferramenta de comparação global. Convidamos desde já outras pessoas a ensaiar outras fórmulas de medida. Talvez os resultados não divirjam muito dos aqui apresentados.

Assim, também poderíamos elaborar um outro gráfico final em que, em vez de apresentar, em forma de percentagens, os quantitativos dos diferentes gráficos parciais, a solução passaria por atribuir 7 pontos a um dicionário cada vez que este apresentasse os melhores resultados em cada uma das variáveis. Ao dicionário que estiver a seguir ser-lhe-iam atribuídos 6 pontos. Ao seguinte, 5 pontos, etc.

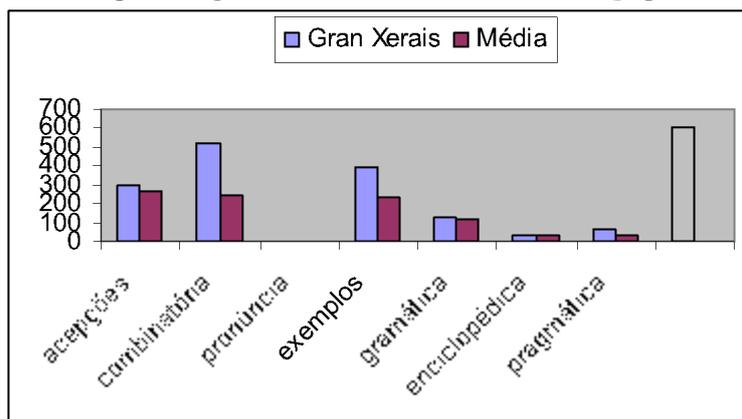
Seguindo este critério de seriação, o dicionário da RAG, por exemplo, deixa de ser favorecido pelo facto de se destacar muito na quantidade de informação fornecida numa das variáveis analisadas (nomeadamente, a dos exemplos). Segundo este critério, em que não são considerados as quantidades de informação fornecida por cada dicionário, mas apenas o número de vezes que ocupam o primeiro lugar, o segundo, etc., os 2 primeiros e os 2 últimos dicionários conservam a sua posição no gráfico final, o *Sotelo Blanco* e o *Gran Cumio* sobem 1 lugar e o dicionário da *RAG* desce 2 posições:



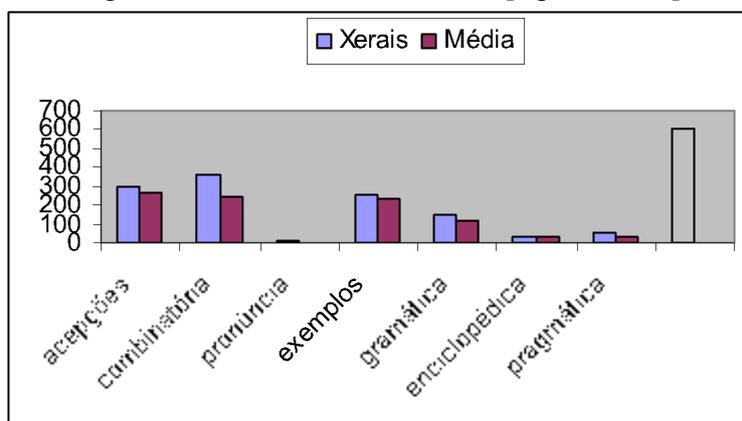
5.2. Os Dicionários monolíngues de que dispomos

Apresentamos a seguir os mesmos dados mas agora não ordenados por variáveis mas sim por dicionários, para, desta maneira, poder apreciar em que se destaca (positiva ou negativamente) cada dicionário analisado.

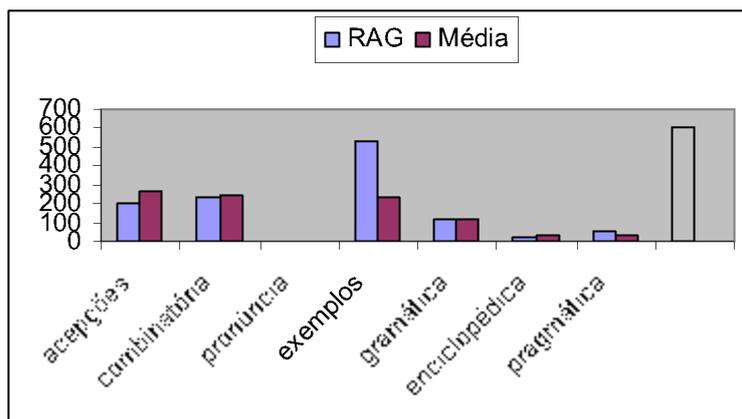
5.2.2. CARBALLEIRA ANLLO, X. M. (coord.) (2000) *Gran Dicionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Gran Xerais*]:



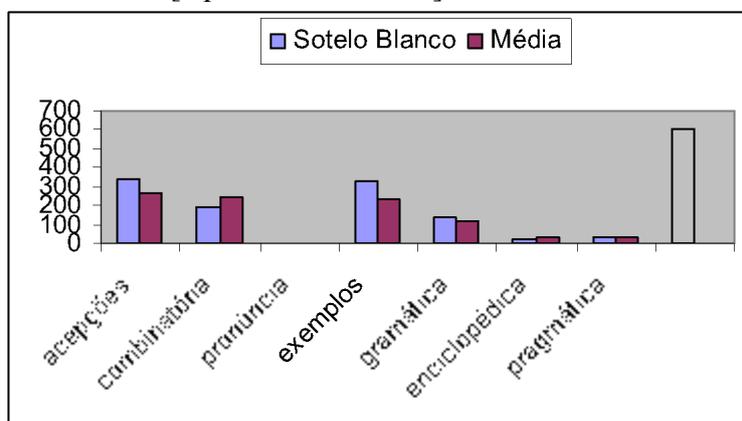
5.2.3. CARBALLEIRA ANLLO, X. M. (coord.) (2004) *Dicionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Xerais*]:



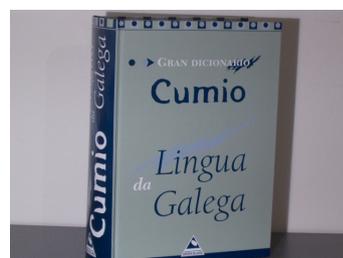
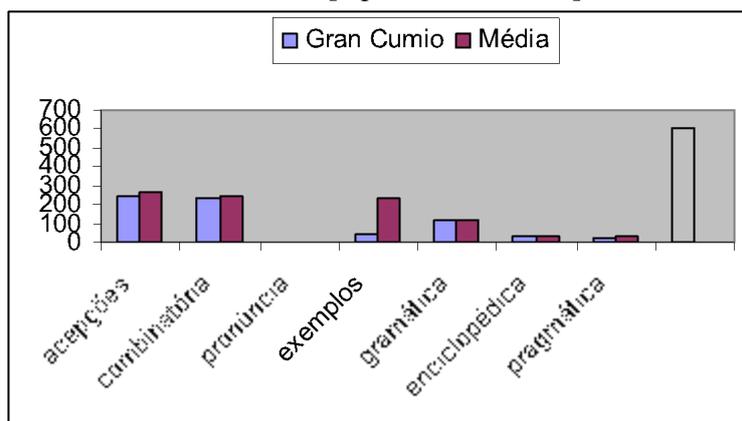
5.2.6. GARCÍA, C. e M. GONZÁLEZ GONZÁLEZ (dir.) (1997) *Diccionario da Real Academia Galega*. Vigo: RAG [aqui: *RAG*]:



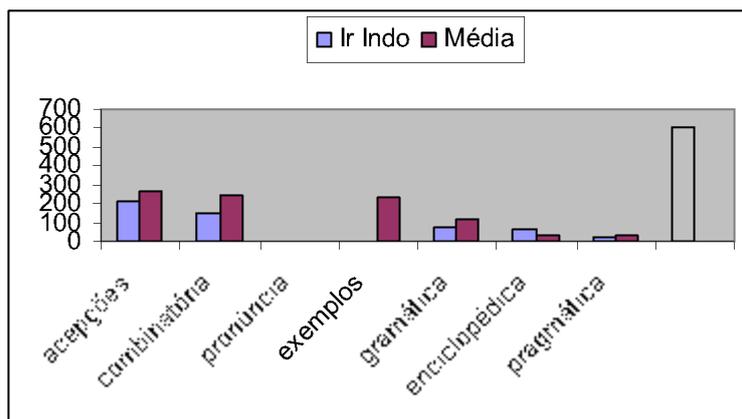
5.2.4. ALONSO ESTRAVÍS, I. (1995) *Diccionario da Língua Galega*. Santiago: Sotelo Blanco [aquí: *Sotelo Blanco*]:



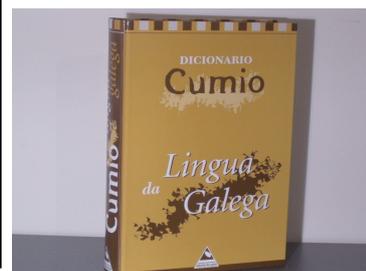
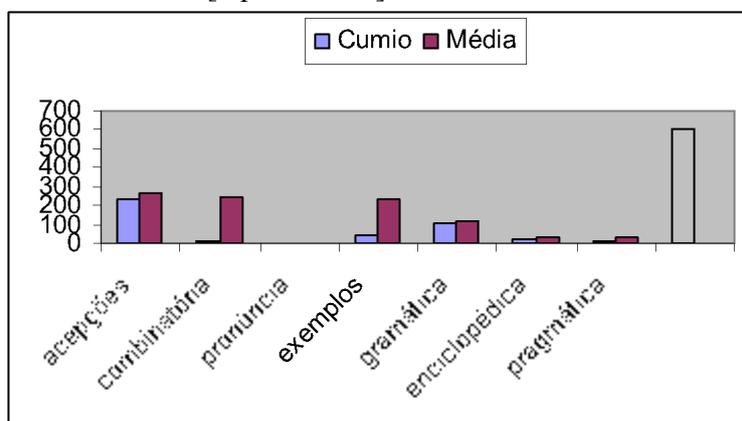
5.2.1. PENA X.A. (dir.) (2004) *Gran Diccionario Cumio da Lingua Galega*. Vigo: Edicions do Cumio [aquí: *Gran Cumio*]:



5.2.3. LEDO CABIDO, B. (dir.) (2004) *Diccionario de Galego*. Vigo: Ir Indo Edicións [aquí: *Ir Indo*]:



5.2.5. PENA X.A. (dir.) (2004) *Diccionario Cumio da Lingua Galega*. Vigo: Edicións do Cumio [aqui: *Cumio*].



3. Referências Bibliográficas

3.1. Dicionários

ALONSO ESTRAVÍS, I.

(1995) *Dicionário da Língua Galega*. Santiago: Sotelo Blanco [aqui: *Sotelo Blanco*].

CARBALLEIRA ANLLO, X. M. (coord.)

(2000) *Gran Dicionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Gran Xerais*].

CARBALLEIRA ANLLO, X. M. (coord.)

(2004) *Diccionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Xerais*].

GARCÍA, C. e M. GONZÁLEZ GONZÁLEZ (dir.)

(1997) *Diccionario da Real Academia Galega*. Vigo: RAG [aqui: *RAG*].

LEDO CABIDO, B. (dir.)

(2004) *Dicionario de Galego*. Vigo: Ir Indo Edicións [aquí: *Ir Indo*].

PENA X.A. (dir.)

(2004) *Dicionario Cumio da Lingua Galega*. Vigo: Edicions do Cumio [aquí: *Cumio*].

PENA X.A. (dir.)

(2004) *Gran Dicionario Cumio da Lingua Galega*. Vigo: Edicions do Cumio [aquí: *Gran Cumio*].

3.2. *Bibliografía geral*

ALONSO RAMOS, M.

(1993) *Las Funciones Léxicas en el modelo lexicográfico de I. Mel'čuk* (tese de doutoramento). Madrid: UNED.

BESSE, H. & R. PORQUIER

(1984) *Grammaires et Didactique des Langues*. Paris: Hatier.

CALDERÓN CAMPOS, M.

(1994) *Sobre la elaboración de diccionarios monolingües de producción. Las definiciones, los ejemplos y las colocaciones léxicas*. Granada: Universidad de Granada.

CARVALHO, J. Herculano de

(1979) *Teoria da Linguagem. Natureza do Fenómeno Lingüístico e a Análise das Línguas*. Vols. 1 e 2. Coimbra: Atlântida.

CATACH, N.

(1981) *Orthographie et lexicographie. Les mots composés*. Paris: Nathan.

COWIE, A. P.

(1983) "On Specifying Grammar. On Specifying Grammatical Form and Function", em Hartmann (ed.) (1983), 99-107.

FRAWLEY, W.

(1981) "In Defense of the Dictionary: A Response to Haiman", em *Lingua. International Review of General Linguistics*, 55-1 (1981), 53-61.

GONZÁLEZ SEOANE, E.

(1996) "A Lexicografía Galega Moderna", em Montegudo e Bouzada Fernández (eds.) (2003), 165-225.

GONZÁLEZ, L.

(2003) "El DRAE en CD-ROM y los millardos", em *Puntoycoma. Boletín electrónico editado por los traductores españoles de la Comisión Europea*, 41 (1996), s.v. **reseñas**.

HAENSCH, G., L. WOLF, S. ETTINGER & R. WERNER

(1982) *La lexicografía. De la lingüística teórica a la lexicografía práctica*. Madrid: Gredos.

HAIMAN, J.

(1980) "Dictionaries and Encyclopedias", em *Lingua. International Review of General Linguistics*, 50-4 (1980), 329-357.

- HARTMANN, R.R.K.
 (1983) “On Specifying Context. How to Label Contexts and Varieties of Usage”, em Hartmann (ed.) (1983), 109-119.
- HARTMANN, R.R.K.(ed.)
 (1983) *Lexicography: Principles and Practice*. London: Academic Press.
- HOUAISS A. (coord.)
 (2001) *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objectiva.
- IRIARTE SANROMÁN, A.
 (2001) *A Unidade Lexicográfica. Palavras, Colocações, Frasemas, Pragmatemas*. Braga: Centro de Estudos Humanísticos-Universidade do Minho.
 (2003) “A informação sobre a categoria gramatical nos dicionários bilingues”, em *Diacrítica – Ciências da Linguagem*, 17-1 (2003), 319-327.
 (2004) “Dicionários Codificadores”, em Sousa, C. M. de e R. Patrício (eds.) (2004), 81-98.
- LEWANDOWSKI, Th.
 (1986) *Diccionario de lingüística*. Madrid: Cátedra.
- LYONS, J.
 (1980) *Semántica*. Barcelona: Teide. [1977].
- MATHIEU-COLAS, M.
 (1994) *Les mots à traits d'union. Problèmes de lexicographie informatique*. Paris: Didier.
- MONTEAGUDO, H. e X.M. BOUZADA FERNÁNDEZ (eds.)
 (2003) *O Processo de Normalización do Idioma Galego (1980-2000)*. Santiago de Compostela: Consello da Cultura Galega-Sección de Lingua.
- SILVA, A. S. da
 (1997) *A Semântica de DEIXAR. Uma Contribuição para a Abordagem Cognitiva em Semântica Lexical*. (Tese de doutoramento. Braga: Universidade Católica Portuguesa-Faculdade de Filosofia de Braga.
- SOUSA, C. M. de e R. PATRÍCIO (eds.)
 (2004) *Largo Mundo Alumiado. Estudos em Homenagem a Vitor Aguiar e Silva*. Braga: Centro de Estudos Humanísticos - Universidade do Minho.
- WERNER, R.
 (1982) “La unidad léxica y el lem”, em Haensch *et al.* (1982), 188-232.
- XAVIER, M. F. e MATEUS, M. H. (eds.)
 (s.d.) *Dicionário de Termos Linguísticos*, I. (Associação Portuguesa de Linguística & Instituto de Linguística Teórica e Computacional). Lisboa: Cosmos.
 (1992) *Dicionário de Termos Linguísticos*, II. (Associação Portuguesa de Linguística & Instituto de Linguística Teórica e Computacional). Lisboa: Cosmos