



---

## Dicionários Monolíngues da Língua Galega\*

---

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.

### Palabras 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.

---

\* Um trabalho semelhante, relativo aos dicionários portugueses, intitulado “Dicionários Codificadores”, foi publicado em Sousa / Patrício (2004: 81-98).

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 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.)<sup>1</sup> suficiente para servirem como ferramentas para ajudar o utilizador a elaborar<sup>2</sup> 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

<sup>1</sup> 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, eclesíásticas, 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.

<sup>2</sup> 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

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 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)<sup>3</sup>.

- a) 15 palavras lexicais ou plenas<sup>4</sup>, das quais:
- Os 7 primeiros substantivos<sup>5</sup>: *ano, vez, día, vida, home, parte, casa*.
  - 7 verbos<sup>6</sup>: *poder, ir, facer, dicir, saber, querer, deber*.
  - 4 adjectivos<sup>7</sup>: *primeiro, novo, mellor, maior*.
  - 2 advérbios: *cando, ben*<sup>8</sup>.

<sup>3</sup> 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, véxase <http://corpus.cirp.es/corga.info.html>

<sup>4</sup> 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 (véxase Lewandowski 1986: s.v. **palabras funcionales**; Xavier e Mateus 1991: s.v. **palavra funcional**).

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

<sup>6</sup> Excluímos *ser, estar, ter* ou *haber*, também com altos índices de frequência, porque consideramos que pertencem ao grupo das palavras funcionais (véxase 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.

<sup>7</sup> 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.). Excluímos *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.

<sup>8</sup> 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.*).

b) 5 palabras gramaticais ou funcionais<sup>9</sup>:– *de, que, e, o, ser.*

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

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)<sup>11</sup>:

<sup>9</sup> Das numerosas palabras gramaticais que encabeçam a lista dos 1000 lemas mais frecuentes 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 esta representado pelo lema *o*) assim como *do* (*de+o*) e *da* (*de+a*), que, de alguma maneira, já están 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 sería o seguinte a ser incorporado, para além de representar lexicograficamente as formas *o*, *a*, *os*, *as* ). Incluímos tamém o lema *ser*, que, como *forma de citação* (Lyons 1980: 20) das várias formas de palabra (*son*, *é*, *era*, *foi*, etc.) registadas entre as 1000 formas mais frecuentes, estaria situada em 6º lugar em termos de frequência no CORGA:

Orde	Forma	Frecuencia	Porcentaxe	Porcentaxe acumulado
1	de	733000	4.154	4.154
2	que	667031	3.780	7.934
3	a	627791	3.558	11.490
4	e	529841	3.002	14.490
5	o	485749	2.753	17.250
5b	ser	± 287264		

<sup>10</sup> Independentemente dos vários critérios de diferenciação que se possam usar (etimológicos, semânticos, morfo-sintáticos), 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<sup>1</sup>** (*v.*) e **poder<sup>2</sup>** (*s.*); **saber<sup>1</sup>** (*v.*) e **saber<sup>2</sup>** (*s.*); **ser<sup>1</sup>** (*v.*) e **ser<sup>2</sup>** (*s.*), etc.

<sup>11</sup> 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.

### 1.4.1. Dicionários Gerais<sup>12</sup>:

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

Carballeira Anllo, X. M. (coord.) (2000): *Gran Diccionario Xerais da Lingua* (Vigo: Edicións Xerais de Galicia) [aqui: *Gran Xerais*].

Ledo Cabido, B. (dir.) (2004): *Diccionario de Galego* (Vigo: Ir Indo) [aqui: *Ir Indo*].

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

### 1.4.2. Dicionários manuais:

García, C. / González González, M. (dir.) (1997): *Diccionario da Real Academia Galega* (A Coruña: RAG) [aqui: *RAG*].

Pena X. A. (dir.) (2004): *Diccionario Cumio da Lingua Galega* (Vigo: Edicións 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.).

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

- Conhecer os seus significados ou aceções, 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 etiquetada<sup>13</sup> num dicionário, melhor será a obra lexicográfica.

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

- (1) Definições e aceções 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.).
- (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 contextual (sintático ou combinatório) e contextual (pragmático ou enciclopédico) de uma determinada unidade lexicográfica. Outro exemplo disto são etiquetas como

---

<sup>13</sup> 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.

FIG. OU POR EXTENSÃO: são etiquetas com informação gramatical (no sentido lato em que aqui entendemos gramatical) ou com informação pragmática-retórica 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)<sup>14</sup>.

Por outro lado, mais importante do que o número de acepções será o tratamento e o desenvolvimento que é dado 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<sup>15</sup>.

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

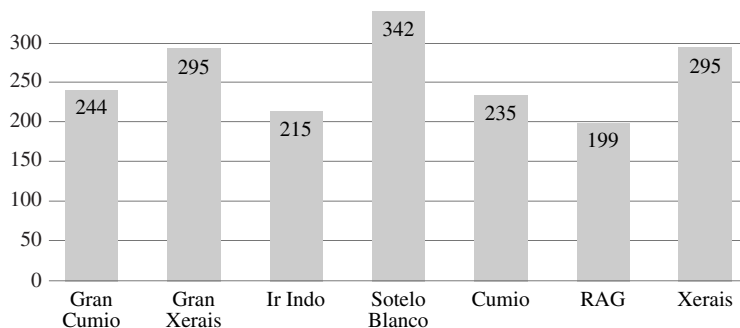


Gráfico 1. Acepções

## 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

<sup>14</sup> 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.

<sup>15</sup> 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 (2001: *passim*)).

como lema da unidade “palavra”<sup>16</sup>, nomeadamente nos dicionários tradicionais em formato não electrónico<sup>17</sup>, onde qualquer tipo de unidade pluriverbal deverá ser registada, em forma de subentrada, sob uma ou várias entradas<sup>18</sup> 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 combinatoria 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

“... **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 combinatoria 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 pluriverbais<sup>19</sup>. Como podemos apreciar no gráfico, o *Gran Xerais* destaca-se no tratamento dado à parte sintagmática:

<sup>16</sup> 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.

<sup>17</sup> 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.

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

<sup>19</sup> 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,



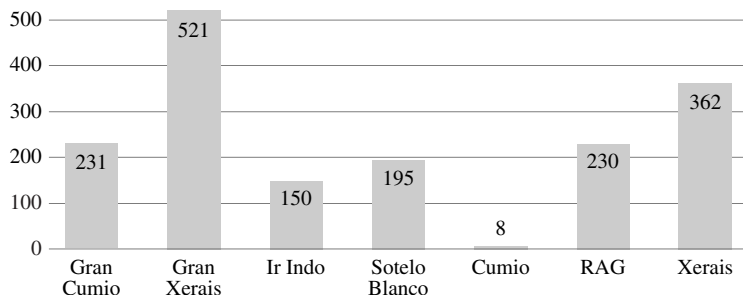


Gráfico 2. Subentradas

### 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'laðo / te'jaðo] *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:

**gato** [ˈgato / ˈɣato] (DIALECTAL, POP.) ...

**dez** [ˈdɛθ / ˈdɛs] (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 metafonía com alternância vocálica em formas como **novoo** vs. **nova**, etc.

---

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, véxase Mathieu-Colas (1994), Catach (1981), Herculano de Carvalho (1979: 506-507, nota 9).

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:

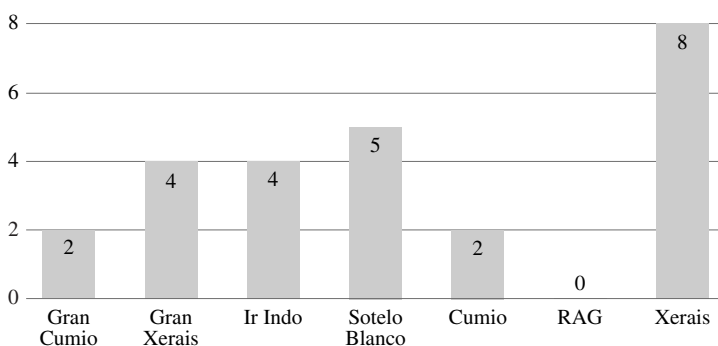


Gráfico 3. Pronúncia

## 2.4. Exemplos e abonações

Não temos grandes dúbidas de que os exemplos e as abonações podem ser muito ricos em información morfológica, sintáctica, combinatoria, semântica, enciclopédica, pragmática, estilística, etc. Os exemplos são cruciais no uso do dicionário como ferramenta para a codificación. 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 sastre” 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” lexicógrafo<sup>20</sup>.

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 edificio*, *fazer uma piscina*, *fazer um estádio*, por um lado, e *fazer un edificio*, *una piscina*, *un 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 der exemplos recolhidos:

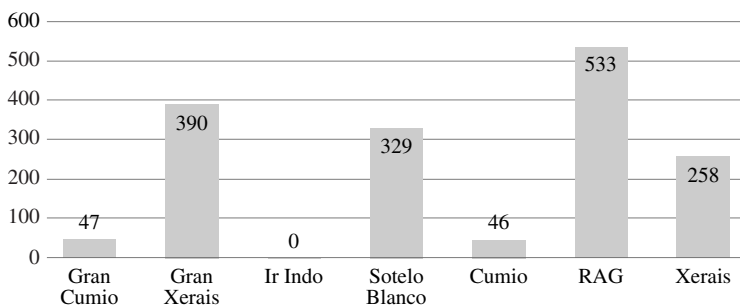


Gráfico 4. Exemplos

## 2.5. Informação gramatical

Analisámos 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:

<sup>20</sup> 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 milieus parisiens, Arnauld et Lancelot (auteurs de la grammaire dite de Port-Royal) illustrent 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).

- A información sobre a categoría (ou subcategoría) gramatical do lema<sup>21</sup>;
- Información ortográfica sobre as unidades lexicográficas;
- Información morfolóxica, como, por exemplo a formación do plural ou do feminino, modelos de conjugación verbal, etc.;
- Información sobre a regêncía verbal (*falar em / falar de*, etc.);
- etc<sup>22</sup>.

A seguir presentamos un gráfico onde tentamos quantificar a información gramatical recollida na totalidade dos lemas estudados:

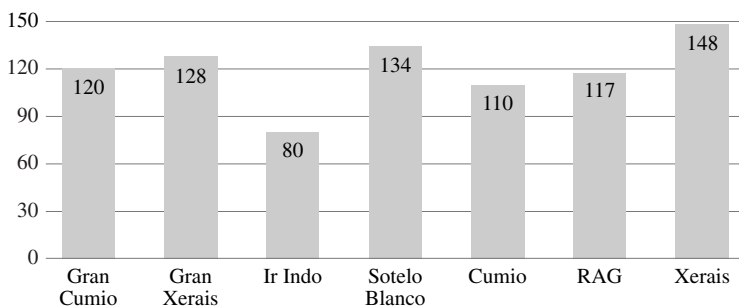


Gráfico 5. Información gramatical

## 2.6. Información enciclopédico-cognitiva

Analizamos aquí a información relativa ao campo ou sistema conceptual, área de coñecemento e ao marco de referencia.

Nos dicionarios actuais, este tipo de información parte moitas veces da intuición do propio lexicógrafo mais do que de unha clasificación produto de unha sistematización

<sup>21</sup> A indicación da categoría gramatical do lema é unha presenza nas obras lexicográficas a que os dicionarios existentes nos teñen habituados e que a maior parte dos autores considera imprescindible. Noutra lugar já falámos (Iriarte Sanromán 2003) sobre a pouca utilidade que as categorizacións gramaticais poden vir a ter para os utilizadores, en grande parte desconhecedores deste tipo de terminoloxía (para alén de que moito frecuentemente os sistemas de etiquetagem actualmente existentes nos dicionarios são excessivamente reductores para os propios linguistas: un verbo transitivo pode intransitivizar-se, por exemplo, ou un verbo que rege preposición pode non ser considerado como sendo transitivo, etc.). Por razóns semellantes, non foran contabilizadas aquí etiquetas como *loc.*, *loc. adv.*, etc.

<sup>22</sup> Decidimos non contabilizar sinónimos e antónimos porque algúns dicionarios fornecen este tipo de información claramente destacadas, con etiquetas como *SIN.* ou *ANT.*, mas outros dicionarios incorporáronos nas propias definicións.

rigorosa<sup>23</sup>. 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 linguística<sup>24</sup>), 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):

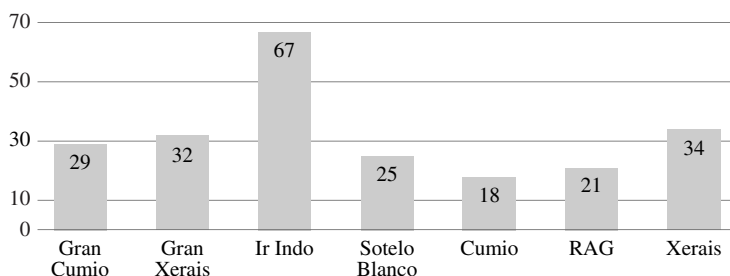


Gráfico 6. Informação enciclopédico-cognitiva

<sup>23</sup> 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 [...]”.

<sup>24</sup> 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, véxase Haiman (1980) e Frawley (1981).

## 2.7. Información pragmático-contextual e retórica

Contabilizámos aquí informacións como:

- Actos de fala: a función ou intención retórico-comunicativa (pense-se por exemplo nas etiquetas do tipo “*usado para*”, “*emprégase como fórmula de cortesía*”, etc.). É fundamental registrar no dicionário de língua os usos de una determinada construción, por exemplo: usado para cumprimentar, despedir-se, ofrecer, convidar, pedir, rexeitar, agradecer; para pedir ou dar informacións, para iniciar una conversa ou establecer turnos de palabra; para pedir consello, dar una orde, chamar a atención, pedir autorización para, negar-se a, ameazar; para exprimir sentimentos, gostos, opinións, actitudes, etc.
- Información sobre variacións formais deliberadas (estruturais ou textuais), etiquetada na tradición lexicográfica con etiquetas como: COLOQ., POÉTICO, FORMAL, etc.
- Información relativa ás variacións sócio-lingüísticas e etno-lingüísticas, etiquetadas na tradición lexicográfica normalmente con abreviaturas como: CAL., FAM., INF., POP.
- Variacións diacrónicas, registadas nos dicionários con etiquetas como ARC., ANTIGO, NEOLOG.
- etc.

Como acontece con a información de tipo enciclopédico-cognitivo, este tipo de información parte moitas veces da intuición do propio lexicógrafo, visto que é moito difícil delimitar, por exemplo, o que é un uso familiar, popular ou figurado<sup>25</sup>. Há, todavía, tentativas de clasificación sistematizada dos diferentes tipos de etiquetas lexicográficas correspondentes aos diferentes contextos e variantes de uso, como a de Hartmann (1983), que establece variantes correspondentes a nove contextos de uso diferentes. O propio autor reconece que a distinción entre un tipo e outro moitas veces é difícil de establecer.

Neste último gráfico antes das conclusións, contabilizámos este tipo de información pragmática nos dicionários analisados:

---

<sup>25</sup> Moitas veces, os adxectivos ou verbos que, xuntamente con un substantivo, conforman una colocación son apresentados lexicograficamente como sendo acepcións ou sentidos figurados (ou impróprios, ou translaticios). Assim, geralmente son consideradas como variacións ou manifestacións de un único sentido figurado ou metafórico acepcións cujo valor só é actualizado quando combinado con outras palabras. É o caso de numerosas colocacións (véxase Alonso Ramos 1993: 169).

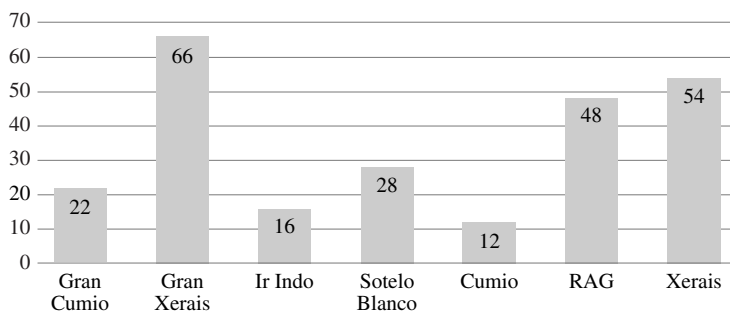


Gráfico 7. Informação pragmático-contextual e retórica

## 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 dicionários<sup>26</sup>.

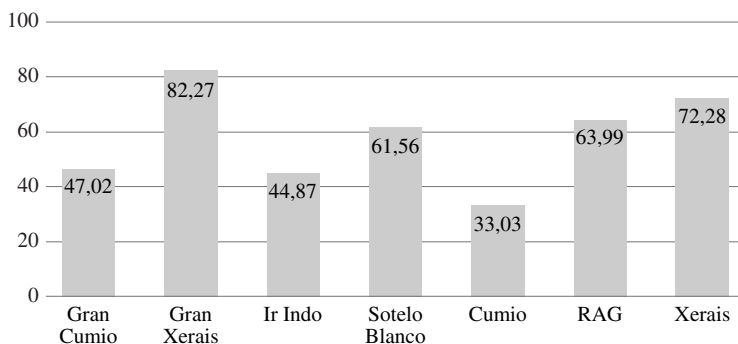


Gráfico 8. Totais

<sup>26</sup> 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 final com os valores totais. Evidentemente, os eventuais erros ou faltas de exactidão são da minha inteira responsabilidade.

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

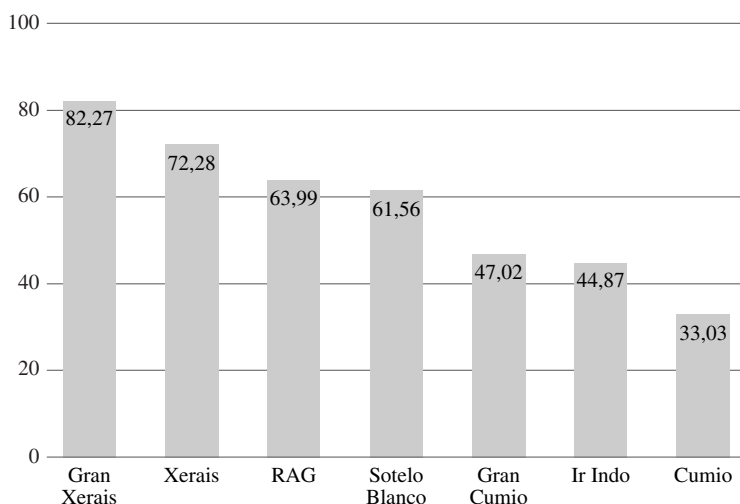


Gráfico 9. “O melhor dicionário é...”

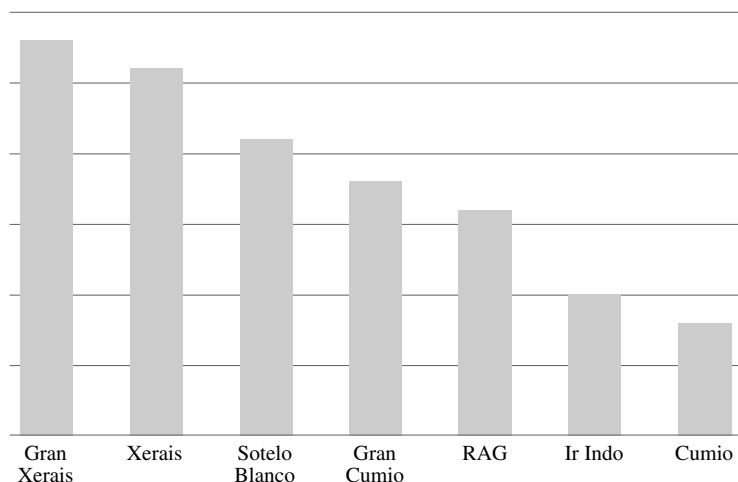
Evidentemente, e como já advertíamos no início deste traballo, este resultado final será sempre cuestionável (a própria quantificación dos dados já é cuestionável). Contudo, pensamos que o resultado pode ser uma boa ferramenta de comparación global. Convidamos desde já outras persoas a ensaiar outras fórmulas de medida. Talvez os resultados non divirjam moito dos aquí presentados.

Assim, tamén poderíamos elaborar un outro gráfico final em que, em vez de presentar, em forma de percentagens, os quantitativos dos diferentes gráficos parciais, a solución pasaría por atribuir 7 puntos a un dicionário cada vez que este presentasse os mellores resultados em cada una das variáveis. Ao dicionário que estiver a seguir ser-lhe-iam atribuídos 6 puntos. Ao seguinte, 5 puntos, etc.

Seguindo este criterio de seriação, o dicionário da RAG, por exemplo, deixa de ser favorecido polo facto de se destacar moito na cantidade de información fornecida numa das variáveis analizadas (nomeadamente, a dos exemplos). Seguindo este criterio, em que non son considerados as cantidades de información fornecida por cada dicionário, mas apenas o número de veces que ocupan o primeiro lugar, o



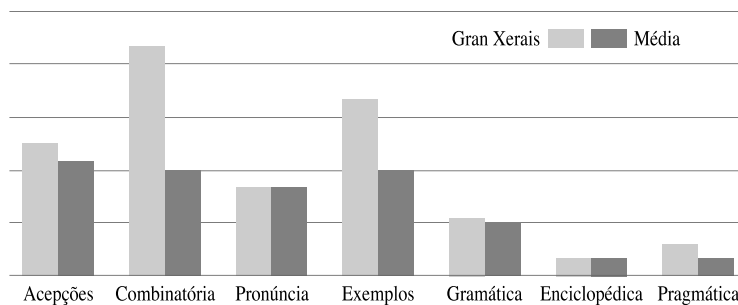
segundo, etc., os 2 primeiros e os 2 últimos dicionários conservam a sua posição no gráfico final, o *Sotelo Blanco* e o *Gran Cumio* sobem 1 lugar e o dicionário da RAG desce 2 posições:



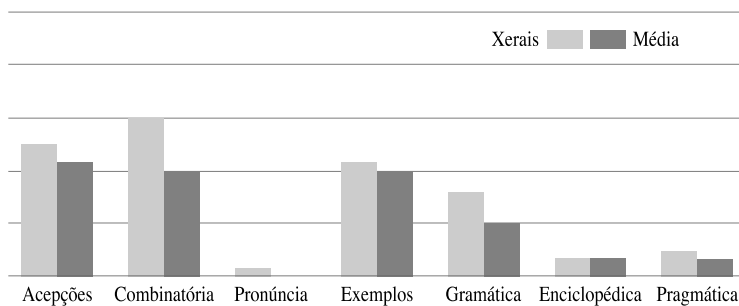
## 5.2. Os Dicionários monolíngues de que dispomos

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.

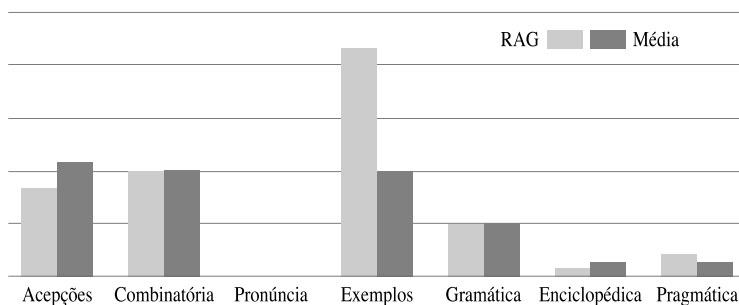
**5.2.1.** Carballeira Anllo, X. M. (coord.) (2000): *Gran Diccionario Xerais da Lingua*. Vigo: Edicións Xerais de Galicia [aqui: *Gran Xerais*]:



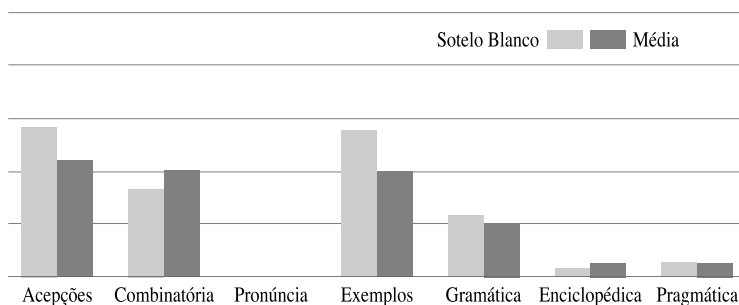
**5.2.2.** Carballeira Anllo, X. M. (coord.) (2004): *Dicionario Xerais da Lingua* (Vigo: Edicións Xerais de Galicia) [aqui: *Xerais*]:



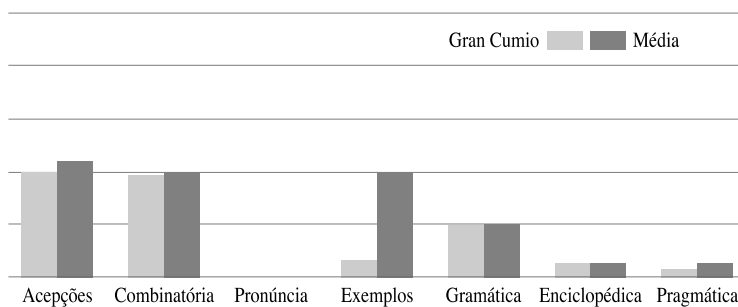
**5.2.3.** García, C. / González González, M. (dirs.) (1997): *Diccionario da Real Academia Galega* (A Coruña: RAG) [aqui: *RAG*]:



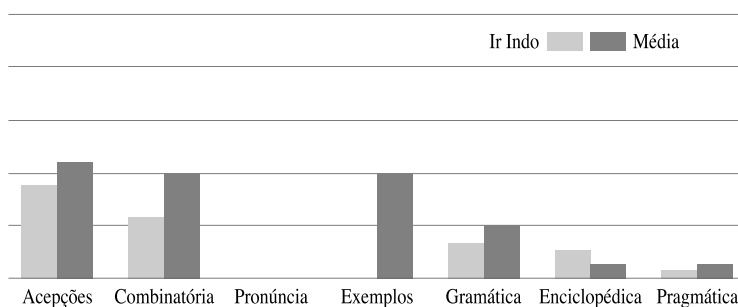
**5.2.4.** Alonso Estravís, I. (1995): *Dicionário da Língua Galega* (Santiago: Sotelo Blanco) [aqui: *Sotelo Blanco*]:



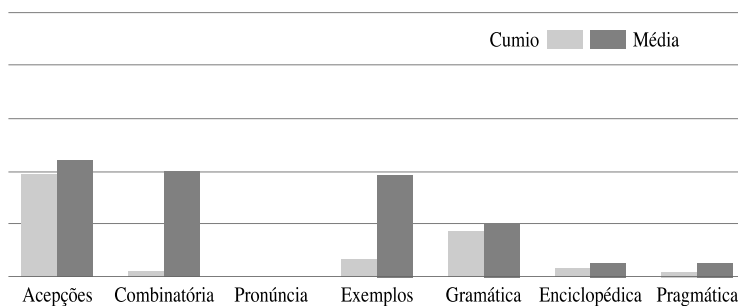
**5.2.5.** Pena X. A. (dir.) (2004): *Gran Dicionario Cumio da Lingua Galega* (Vigo: Edicións do Cumio) [aqui: *Gran Cumio*]:



**5.2.6.** Ledo Cabido, B. (dir.) (2004): *Diccionario de Galego* (Vigo: Ir Indo) [aqui: *Ir Indo*]:



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



## Referências Bibliográficas

### 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): *Dicionario Xerais da Lingua* (Vigo: Edicións Xerais de Galicia) [aqui: *Xerais*].
- García, C. / González González, M. (dirs.) (1997): *Diccionario da Real Academia Galega* (A Coruña: RAG) [aqui: RAG].
- Ledo Cabido, B. (dir.) (2004): *Dicionario de Galego* (Vigo: Ir Indo) [aqui: *Ir Indo*].
- Pena X. A. (dir.) (2004): *Dicionario Cumio da Lingua Galega* (Vigo: Edicións do Cumio) [aqui: *Cumio*].
- Pena X. A. (dir.) (2004): *Gran Dicionario Cumio da Lingua Galega* (Vigo: Edicións do Cumio) [aqui: *Gran Cumio*].

### Bibliografía geral

- Alonso Ramos, M. (1993): *Las Funciones Léxicas en el modelo lexicográfico de I. Mel' iuk*. Tese de Doutoramento (Madrid: UNED).
- Besse, H. / Porquier, R. (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, R. R. K. (ed.), *Lexicography: Principles and Practice*: 99-107 (London: Academic Press).
- Frawley, W. (1981): "In Defense of the Dictionary: A Response to Haiman", *Lingua. International Review of General Linguistics*, 55-1: 53-61.
- González Seoane, E. (2003): "A Lexicografía Galega Moderna", em Monteagudo, H. / Bouzada Fernández, X. M. (eds.), *O Proceso de Normalización do Idioma Galego (1980-2000)*: 165-227 (Santiago: Consello da Cultura Galega-Sección de Lingua).

- 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. / Wolf, L. / Ettinger, S. / Werner, R. (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: 329-357.
- Hartmann, R. R. K. (1983): “On Specifying Context. How to Label Contexts and Varieties of Usage”, em Hartmann, R. R. K. (ed.), *Lexicography: Principles and Practice*: 109-119 (London: Academic Press).
- 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, Frasesmas, Pragmatemas* (Braga: Centro de Estudos Humanísticos – Universidade do Minho).
- Iriarte Sanromán, A. (2003): “A informação sobre a categoria gramatical nos dicionários bilíngues”, *Diacrítica – Ciências da Linguagem*, 17-1: 319-327.
- Iriarte Sanromán, A. (2004): “Dicionários Codificadores”, em Sousa, C. M. de / Patrício, R. (eds.), *Largo Mundo Alumiado. Estudos em Homenagem a Vítor Aguiar e Silva*: 81-98 (Braga: Centro de Estudos Humanísticos – Universidade do Minho).
- Lewandowski, Th. (1986): *Diccionario de lingüística* (Madrid: Cátedra).
- Lyons, J. (1980) [1977]: *Semántica* (Barcelona: Teide).
- Mathieu-Colas, M. (1994) : *Les mots à traits d’union. Problèmes de lexicographie informatique* (Paris: Didier).
- Monteagudo, H. / Bouzada Fernández, X. M (eds.) (2003): *O Proceso de Normalización do Idioma Galego (1980-2000)* (Santiago: 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.
- Sousa, C. M. de / Patrício, R. (eds.) (2004): *Largo Mundo Alumiado. Estudos em Homenagem a Vítor Aguiar e Silva* (Braga: Centro de Estudos Humanísticos – Universidade do Minho).

- Werner, R. (1982): “La unidad léxica y el lema”, em Haensch, G. / Wolf, L. / Ettinger, S. / Werner, R. (eds.), *La lexicografía. De la lingüística teórica a la lexicografía práctica*: 188-232 (Madrid: Gredos).
- Xavier, M. F. / Mateus, M. H. (eds.) (s.d.): *Dicionário de Termos Linguísticos*, I. Associação Portuguesa de Linguística / Instituto de Linguística Teórica e Computacional (Lisboa: Cosmos).
- Xavier, M. F. / Mateus, M. H. (eds.) (1992): *Dicionário de Termos Linguísticos*, II. Associação Portuguesa de Linguística / Instituto de Linguística Teórica e Computacional (Lisboa: Cosmos).