

Dicionários Codificadores*

ÁLVARO IRIARTE SANROMÁN

(Universidade do Minho)

1. O objectivo deste trabalho será mostrar em que medida os principais dicionários monolíngues portugueses de que dispomos no mercado fornecem informação (gramatical, combinatória, pragmática, etc.) suficiente para servirem como ferramentas para produzir ou codificar um texto em português. Tentaremos quantificar, de maneira aproximada, esta informação, para assim poder justificar a resposta a perguntas que frequentemente me foram colocadas, do tipo: Qual é o melhor dicionário de português? Que dicionário posso utilizar para aprender português? Que dicionário me pode ajudar a escrever em português?, etc.

O Professor Aguiar e Silva sabe da forte impressão e da crise que provocou em mim a leitura de *Imposturas Intelectuais* (Sokal & Bricmont, 1999). O presente texto, marcado profundamente por essa leitura, é o primeiro que escrevo após a mesma. 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.

2. 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.).

* O Professor Aguiar e Silva (que nas suas aulas de mestrado tanto me fascinou com a profundidade, mas também com a clareza e com a fluência do seu discurso, oral e escrito) levou-me a duvidar sobre a minha clara orientação para a linguística. Felizmente, ou quiçá graças às sátiras de Correia Garção (que tive de analisar no segundo semestre daquele mestrado e que –desculpe, Professor– sempre detestarei), voltei a enveredar pelos caminhos da linguística. Mas mesmo nestes, o Professor Aguiar e Silva foi também o responsável, não me cansarei de o repetir, pelo empurrão definitivo que me levou ao mundo da lexicografia, prática e teórica, ao convidar-me a elaborar para a Porto Editora um diagnóstico do seu dicionário de Espanhol-Português e posteriormente ao convidar-me a formar uma equipa para a elaboração de um novo dicionário.

- Conhecer as suas capacidades combinatórias, as relações possíveis com outras unidades no co-texto (combinações livres, colocações, frases, 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 etiquetada num dicionário, melhor será a obra lexicográfica, nomeadamente se estiver orientada para a codificação linguística.

Um dicionário codificador ou de produção é uma ferramenta pensada para ajudar o utilizador a elaborar textos numa língua, normalmente segunda ou estrangeira, ainda que também se possa conceber um dicionário codificador para a língua materna. A característica mais importante deste tipo de dicionários 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.

A distinção entre dicionários descodificadores, ou de recepção, e dicionários codificadores, ou de produção, está já consolidada na teoria lexicográfica, embora, na prática lexicográfica, na prática editorial, se continue a não fazer tal distinção, publicando-se dicionários teoricamente bidireccionais¹.

Um dicionário codificador pode ser também um dicionário de aprendizagem, mas não devemos identificar dicionários de produção ou dicionários codificadores com dicionários de aprendizagem, didácticos ou pedagógicos. Estes últimos são dicionários com uma clara orientação escolar², que se caracterizam por conter um número mais reduzido de entradas; usarem definições abreviadas e simples; usarem um léxico corrente que facilite a aquisição por parte dos alunos de novas palavras e novos conceitos; terem uma função polivalente (são ao mesmo tempo dicionários de

¹ Sobre dicionários codificadores, e dicionários de uso, *vd.*, entre outros, Al-Kasimi (1983), Benson, M., E. Benson & R. Ilson (1986a; 1986b), Bogaards (1991, 1992), Calderón (1994), Haensch (1982a: 156-158), Hernández (1989), Ilson (ed.) (1985), Lehmann (1991), Martin (1985), Moulin (1983), Sinclair (1991), Hartmann (ed.) (1984), Schafroth (1998).

² Sobre dicionários de orientação escolar, *vd.* Hernández (1989).

sinónimos e antónimos, ortográficos, etc.); terem certa orientação normativa; serem, muito frequentemente, ilustrados, etc. O *Dicionário do Português Básico* (Vilela, 1991) é um bom exemplo de dicionário de aprendizagem.

É importante também não identificar dicionário codificador com dicionário onomasiológico, analógico ou conceptual (vd. Iriarte Sanromán, 2001: 32-35).

3. Para realizarmos a análise dos dicionários, seleccionámos os seguintes lemas, tomados duma listagem dos 100 lemas mais frequentes do *corpus* CETEMPúblico (<http://www.linguateca.pt/CETEMPUBLICO>)³:

a) 15 palavras lexicais ou plenas⁴, das quais:

- Os 5 primeiros substantivos⁵: *ano, dia, país, vez, conto*.
- 5 verbos⁶: *fazer, ir, dizer, passar, dar*.
- 3 adjectivos⁷: *grande, bom, último*,
- 2 advérbios: *ontem, hoje*⁸.

b) 4 palavras gramaticais ou funcionais⁹:

- *de, o, que, ser*.

Serão objecto de análise os grandes dicionários de língua:

AULETE, F. J. Caldas (1987): *Dicionário da Língua Portuguesa Caldas Aulete*. [aqui: *Caldas Aulete*].

³ Agradeço ao Engenheiro José João Almeida o envio da listagem dos 100 lemas mais frequentes do *corpus* CETEMPúblico. Sobre este *corpus*, vd. Rocha & Diana (2000).

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, 1988: s.v. **palabras funcionales**, e Xavier & Mateus (org.), 1991: s.v. **palavra funcional**).

⁵ Excluimos *português*, com índice de frequência maior do que *conto*, devido à sua frequente função de adjectivo.

⁶ Excluimos *haver, ter, estar e ser*, também com altos índices de frequência, porque consideramos que pertencem ao grupo das palavras funcionais.

⁷ Excluimos *segundo* e *só*, com maiores índices de frequência, por terem frequentemente outras funções para além da de adjectivo.

⁸ Em termos de frequência, *ontem* e *hoje* são os dois últimos 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: *não, mais, muito, já, ainda*.

⁹ Das numerosas palavras gramaticais que encabeçam a lista dos 100 lemas mais frequentes do *corpus*, escolhi as primeiras, excluindo a forma *a* (que poderia ser artigo, pronome, substantivo masculino e preposição e que, lexicograficamente é representado pelo lema *o*), o lema *que* (que também pode funcionar como pronome relativo ou interrogativo, como advérbio ou como conjunção), assim como *de+o* e *de+a* (que já estavam incluídos, embora separadamente, nos primeiros lugares da lista). Embora o lema *o* também possa ter as funções de substantivo masculino, artigo definido e pronome, a sua frequência é quatro vezes superior à do lema *para*, que seria o seguinte a ser incorporado, para além de representar lexicograficamente as formas *o, a, os, as*).

- CASTELEIRO, J. Malaca (coord.) (2001): *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*. [aqui: *Academia*].
- FERREIRA, A. Buarque de Holanda (1999): *Novo Aurélio Século XXI: O Dicionário da Língua Portuguesa*. [aqui: *Aurélio*].
- FIGUEIREDO, Cândido de (1982): *Dicionário da Língua Portuguesa*. [aqui: *Cândido*].
- HOUAISS A. (coord.) (2001): *Dicionário Houaiss da Língua Portuguesa*. [aqui: *Houaiss*].
- MACHADO, J. P. (1981): *Grande Dicionário da Língua Portuguesa*. [aqui: *Machado*].
- WEISZFLOG, W. (1998): *Moderno dicionário da Língua Portuguesa*. [aqui: *Michaelis*].
- SILVA, A. Morais da (1980): *Novo Dicionário Compacto da Língua Portuguesa*. [aqui: *Morais*].
- (1996) *Novo Dicionário da Língua Portuguesa. Sistema J*. [aqui: *Lello*].¹⁰

Analisámos também dois dicionários que pertencem claramente a um segmento inferior (o que deverá ser tido em conta na leitura dos gráficos), mas muito conhecidos do grande público e com um uso frequentemente escolar, embora não sejam propriamente dicionários de aprendizagem nem dicionários codificadores:

- ALMEIDA Costa J. e A. SAMPAIO E MELO (1994): *Dicionário da Língua Portuguesa*. [aqui: *Porto Editora*].
- (1995): *Dicionário Universal da Língua Portuguesa*. [aqui: *Universal*].

Estes dois dicionários estão aqui mais como ponto de referência do que para serem comparados com os primeiros.

Embora seja o único dicionário português concebido como dicionário de aprendizagem e também como dicionário de codificador¹¹ excluímos o *Dicionário do Português Básico* (Vilela, 1991) uma vez que é claramente um pequeno dicionário escolar, de boa qualidade mas que não pode ser comparado, em termos quantitativos, com os primeiros.

Nas 19 entradas indicadas pesquisaremos e tentaremos quantificar os seguintes aspectos:

- (1) Número de subentradas [expressões pluriverbais formadas pelo lema mais outra(s) palavra(s)];

¹⁰ Faltam nesta lista os dicionários do Círculo de Leitores. A primeira edição por ser a mesma de *Machado*, a segunda porque não dispomos de dados completos (só existe o 1º volume no Depósito Legal da Bibliote Pública de Braga), a terceira edição, versão europeia do *Houaiss*, por estar ainda incompleta.

¹¹ Sobre o *Vilela* como dicionário codificador, *vd.* Shafroth (1998).

- (2) Número de acepções.
- (3) Transcrição(-ções) fonética(s) ou figurada(s).
- (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).

Excluimos os anexos gramaticais, etc. porque consideramos que a maior parte dos utilizadores "normais" dos dicionários não utilizam este tipo de suplementos na altura da consulta do dicionário.

4. Antes de começarmos a análise dos dicionários, será 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 pragmáticos, 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 (sintático, combinatório) e contextual (pragmático-enciclopédico) de uma determinada unidade lexicográfica. Outro exemplo disto são etiquetas como FIG. ou POR EXTENSÃO: são etiquetas com informação gramatical ou com informação pragmática como COLOQ. VULG. ou DESUS.? Esta dificuldade em separar claramente os diferentes tipos de informação vai ter consequências no cômputo final que estará por trás dos gráficos apresentados, o que nos levará a apresentar os dados quantitativos como sendo sempre aproximados.

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 possível¹², que ultrapasse sem nenhum tipo de receio o imanentismo gramatical a que nos habituou a linguística do século XX.

¹² Neste sentido, não pensamos que um sistema de etiquetagem seja, *a priori*, mais imperfeito ou incompleto do que uma definição. Para além de concebermos a definição como fazendo parte do conjunto de etiquetas, podem existir bons e maus sistemas de etiquetagem e boas (no sentido mais pragmático de 'úteis') e más definições.

Repare-se, por exemplo, na insuficiência das definições de adjectivos como *esquisito* ou *espantoso* em alguns dicionários portugueses. Ao consultarmos nos dicionários as entradas correspondentes à primeira palavra deparamos com acepções deste termo da mais variada ordem, positivas e negativas. Em alguns casos, segundo as definições apresentadas, o termo parece chegar a funcionar como antónimo de si próprio (para além de que parecem ter ficado surpreendentemente desactualizadas uma vez que a maior parte das acepções são definidas em termos positivos).

Temos aqui um caso claro em que o valor de cada acepção da palavra só deixará de ser ambíguo com a informação fornecida pelo co-texto (*iguarias esquisitas*) e/ou pelo contexto situacional (*Deve haver alguma coisa estragada no frigorífico: há um cheiro esquisito*), o que, em termos lexicográficos, implicará a necessidade de um bom sistema de etiquetagem, com boas definições, naturalmente, mas também com informações sobre contextos e restrições de uso, etc.

4.1. Número de subentradas

Qualquer palavra, signo, letra (ou conjunto de palavras, signos ou letras) poderia constituir uma entrada num dicionário. Contudo, no que se refere à combinatoria lexical, as vantagens da unidade palavra¹³ como lema são claras, nomeadamente nos dicionários tradicionais em formato não electrónico¹⁴, onde qualquer tipo de unidade pluriverbal¹⁵ deverá ser registada, em forma de subentrada, sob uma ou várias entradas das várias palavras lexicais que compõe a expressão pluriverbal¹⁶, correspondendo assim ao que Cowie (1983: 99) chama “expectativas conservadoras dos usuários comuns dos dicionários”.

¹³ 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.

¹⁴ 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.

¹⁵ Contudo, o tratamento não será o mesmo para os *frasesmas completos*, para as *colocações*, etc. (vd. Iriarte Sanromán, 2001: § 4.5 e § 5.4).

¹⁶ Werner (1982: 224-229) apresenta vários argumentos para defender que o lema deverá corresponder à unidade palavra:

- a maior parte dos dicionários é elaborada para um público amplo, sem formação nem interesses linguísticos;
- “monema”, “morfema” e “lexema” não são unidades operacionais nem sequer entre profissionais da língua, como intérpretes, tradutores, terminólogos, escritores, etc.
- a unidade linguística “palavra” está bem enraizada na consciência de todos os usuários de dicionários, o que não acontece com termos e conceitos como “monema”, “morfema” e “lexema”;

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). Tais acepções são, de facto, combinações lexicais que deveriam ser registadas em forma de subentradas.

Apresentamos a seguir um gráfico que, de alguma maneira, tenta quantificar o tratamento que cada dicionário dá à importante questão da combinatória lexical uma vez que nas subentradas está a indicar-se explicitamente que o valor (acepção) da palavra vem dado pela combinação da mesma com outras palavras¹⁷.

-
- a unidade linguística “palavra” tem uma grande importância cultural (na arte, na religião, no direito, na publicidade, etc.);
 - no processo de aprendizagem da língua materna (pelo menos nas línguas ocidentais com tradição escrita), a unidade palavra consolida-se associada ao conhecimento e à estruturação conceptual da realidade: quando a criança recebe instruções explícitas sobre o conteúdo referencial dos significantes linguísticos (“Isto é um...”, “Como se chama isto...?”, etc.), estas estão associadas a palavras, não a monemas isolados, habituando-se desta forma a considerar as palavras como nomes de classes de elementos da realidade e de conceitos;
 - no ensino da língua materna e de línguas segundas, a unidade “palavra” continua a desempenhar um papel primordial, apesar dos docentes e dos manuais escolares, ou graças a eles;
 - a importância do código escrito nas comunidades linguísticas ocidentais, e particularmente no ensino institucionalizado de línguas.

¹⁷ Não contabilizaremos as formas pluriverbais com direito a entrada diferenciada nos dicionários, como por exemplo *dia-a-dia*, *conto-do-vigário*, etc. porque iria desvirtuar, em termos quantitativos, o resultado final, uma vez que suporia contabilizar a informação gramatical correspondente (sobre a categoria do lema), a informação fonética, 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 —*cf.*, por exemplo, *primeiro ministro* (Vilela, *s.v.* **salário**) *vs.* *primeiro-ministro* (Vilela, *s.v.* **reunir**); *caminho-de-ferro* (Porto Editora) *vs.* *caminho de ferro* (Cândido) *vs.* *estrada de ferro* (Porto Editora e Aurélio), *co-opositor* *vs.* *coocupante* (Aurélio) *vs.* *co-ocupante* (Porto Editora); etc.—. Sobre o assunto, *vd.* Mathieu-Colas (1994), Catach (1981), Herculano de Carvalho (1979: 506-507, nota 9).

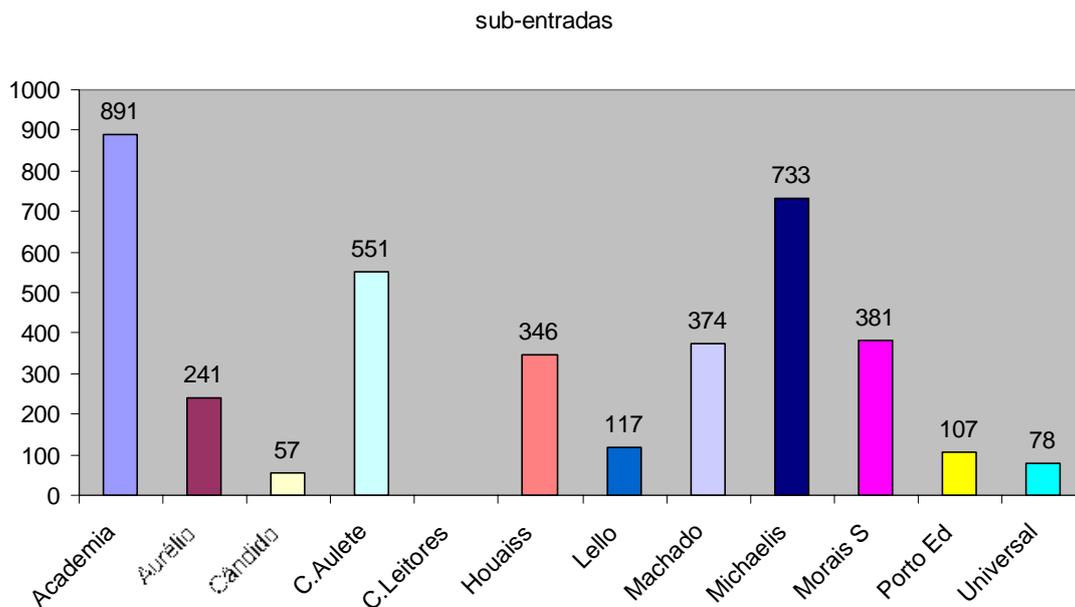


Gráfico 1
Subentradas

4.2. Número de acepções

Nem sempre é fácil delimitar e contabilizar o número de acepções. Por isso serão contabilizadas em termos aproximados.

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.

Não podemos também fazer corresponder as acepções (com definições em forma de sinónimos) separadas por ponto e vírgula num dicionário com as acepções separadas numericamente noutros dicionários. Assim, no *Cândido* ou no *Aurélio*, dentro das acepções separadas numericamente podemos encontrar outras subacepções separadas por ponto e vírgula; e, no *Houaiss*, encontraremos acepções e subacepções separadas por números numa estrutura semelhante à de um índice (1, 2, **2.1.**, **2.2.**, **2.3.**, **2.3.1.**, 2.4., etc.).

Esta impossibilidade ou dificuldade em contabilizar o número de acepções reflecte a natureza não discreta do fenómeno do significado, uma vez que os

significados não são como as coisas, entidades separáveis, contáveis (*cf.* Silva, 1997: 587-588)¹⁸.

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

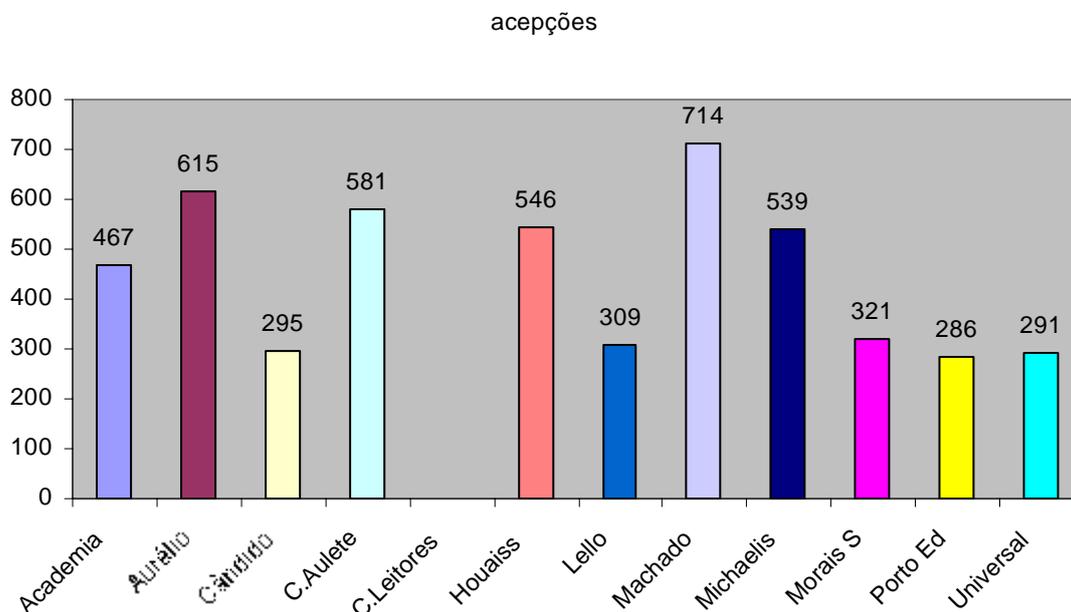


Gráfico 2
Acepções

4.3. Transcrição(-ções) fonética(s) ou figurada(s)

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

carro [ˈkaɾo / ˈkaɾo] *s.m.* 1. ...

ministro [miˈnistɾu / miˈnistɾu] *s.m.* 1. ...

Poderão também registar-se variantes claramente afastadas daquilo que poderíamos considerar português padrão, embora muito usadas e também muito desprezadas ou ignoradas nos manuais de fonética e nas gramáticas. A(s) forma(s) considerada(s) padrão ficariam marcadas pelo facto de não ter nenhum tipo de etiqueta:

vivo [ˈvivu / ˈβiβu (NORTE, POP.)] ...

cama [ˈkama / ˈkama (NORTE)]

chouriço [ʃoˈɾisu / ʃouˈɾisu / tʃouˈɾisu (NORTE, POP.) / ʃoˈɾisu

¹⁸ 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.

(ALENTEJO, POP.)] ...

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 destaca-se claramente o dicionário da Academia, que apresenta uma transcrição fonética completa do lema e da correspondente forma feminina. Outros dicionários apresentam, em forma de pronúncia figurada, o grau de abertura da vogal tónica ou a pronúncia de alguma grafia específica (*cf.*, por exemplo, o [z] na entrada **exame** ou **trânsito** no *Porto Editora*).

Em termos quantitativos atribuímos, para a elaboração do correspondente gráfico, 2 valores à transcrição fonética de cada lema, 1 valor à pronúncia figurada de vogais tónicas, etc., e 1 valor para determinados desenvolvimentos de tipo gramatical sobre alguns aspectos da pronúncia. Para a elaboração dos totais (*vd. infra* gráficos 8 e 9) os valores relativos à informação sobre a pronúncia serão somados aos quantitativos relativos à informação gramatical.

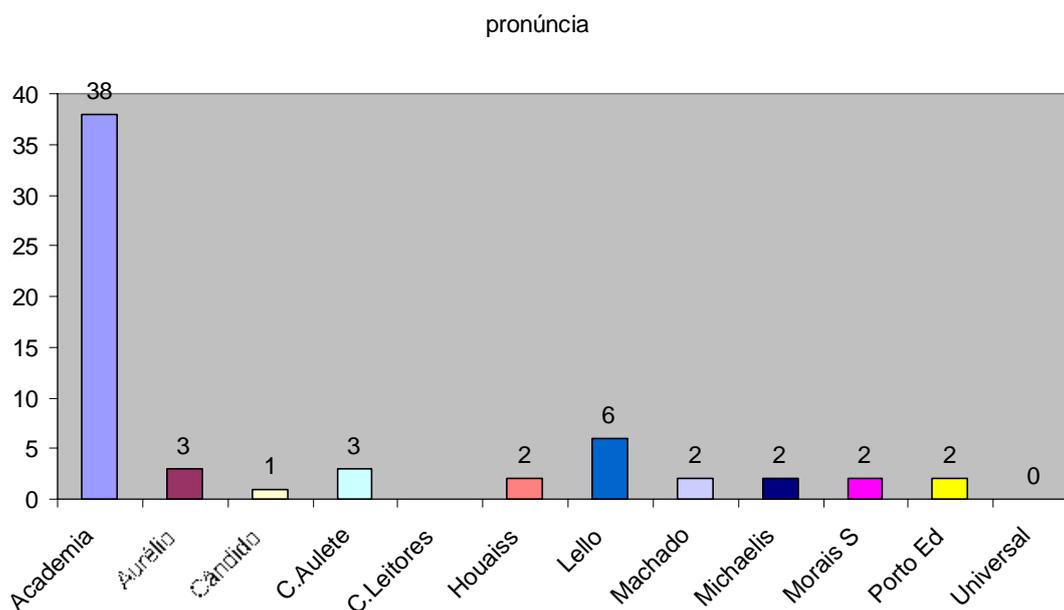


Gráfico 3

Pronúncia

4.4. 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 panaceia ou de “caixote do lixo” 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” lexicógrafo¹⁹.

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. Em primeiro lugar porque nem sempre é fácil distinguir entre o que são exemplos e o que são combinações lexicais (que seriam contabilizadas como subentradas). Exemplos disto podemos encontrá-los no dicionário da Academia. Por outro lado, como é que devemos contabilizar casos como: *fazer um edifício*, *fazer uma piscina*, *fazer um estádio*, por um lado, e *fazer um edifício*, *uma piscina*, *um estádio*, por outro?

¹⁹ 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).

Eis a seguir o gráfico correspondente ao número de exemplos da totalidade dos lemas estudados.

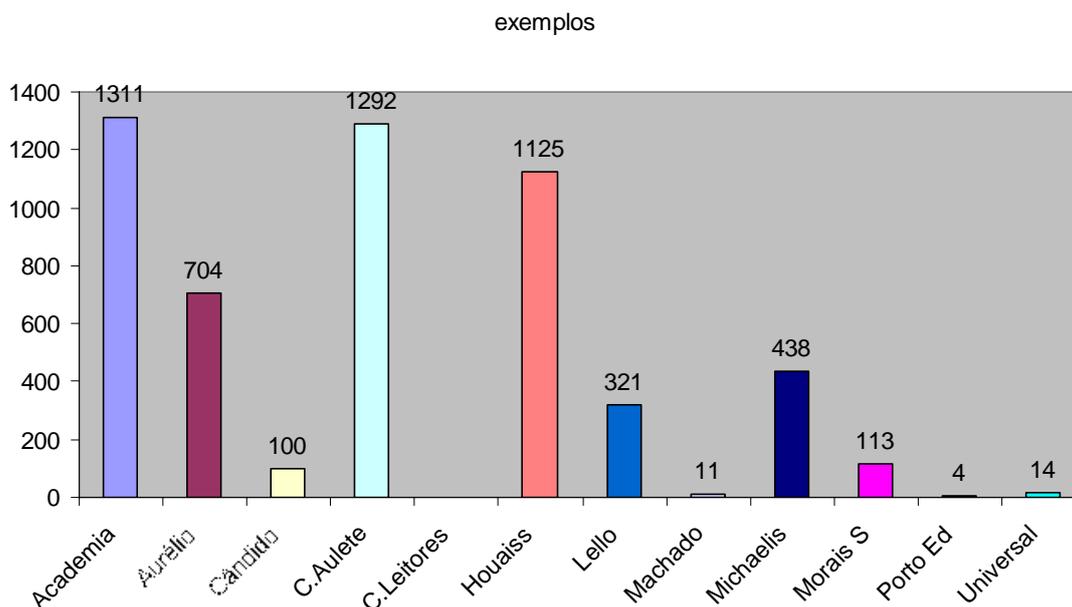


Gráfico 4

Exemplos

4.5. Informação gramatical

Com "etiquetagem gramatical" queremos referir-nos à presença de qualquer tipo de informação gramatical no dicionário, nomeadamente informações ou restrições de carácter ortográfico, morfológico, sintáctico, semântico ou lexical relativa a cada lexema ou combinação de lexemas que conformem uma unidade lexicográfica.

Contabilizámos neste grupo:

- A informação sobre a categoria (ou subcategoria) gramatical do lema²⁰;
- Informação ortográfica sobre as unidades lexicográficas;
- Informação morfológica, como, por exemplo a formação do plural ou do feminino, modelos de conjugação verbal, etc.

²⁰ A indicação da categoria gramatical do lema é uma presença nas obras lexicográficas a que nos têm habituados os dicionários existentes 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 não é considerado transitivo, etc.).

- Restrições à combinação sintáctico-semântica das palavras (como por exemplo, *homem bom/bom homem, vivo interesse, falar em/falar de*, etc.);

- Restrições à combinação lexical dos lexemas (a informação colocacional é muito importante, nomeadamente no caso dos dicionários orientados para a codificação, como por exemplo: *lâmpada fosca, papel mate; talhar-se o leite, avinagrar-se o vinho*, etc.);

- Antónimos, homónimos e parónimos. Não contabilizaremos a informação sobre sinónimos uma vez que alguns dos dicionários utilizam sinónimos nas sua definições.

- etc.

Como já indicámos (*vd. supra* § 4.3), para a elaboração dos gráficos finais juntámos aos totais relativos à informação gramatical os quantitativos relativos à informação sobre a pronúncia.

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

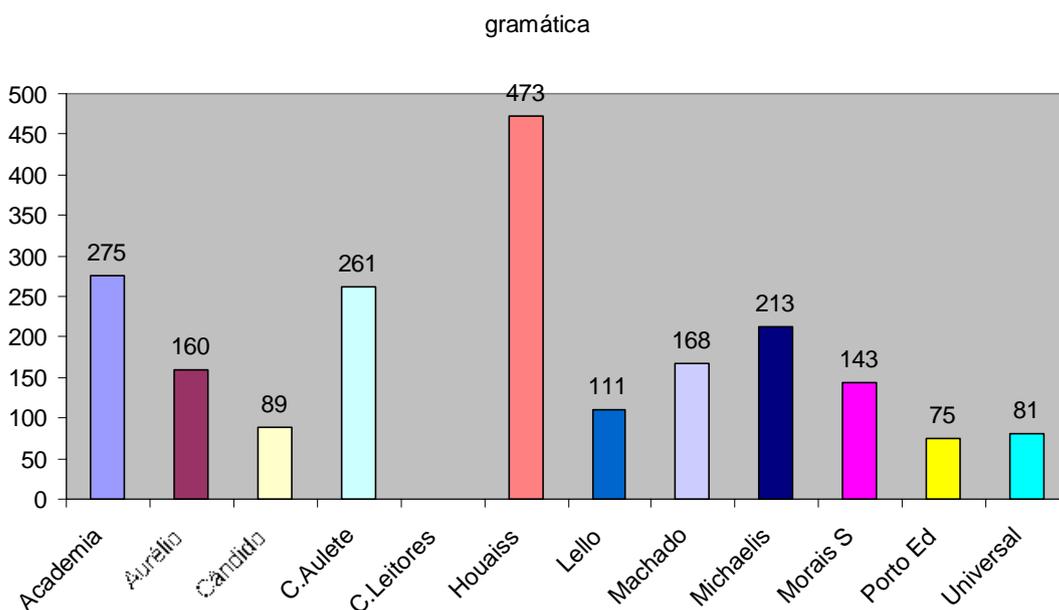


Gráfico 5
Informação gramatical

4.6. Etiquetagem enciclopédico-cognitiva

Com "etiquetagem enciclopédico-cognitiva" referimo-nos à informação relativa ao campo ou sistema conceptual, área de conhecimento e ao marco de

referência. 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 rigorosa²¹. 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 alimentação, 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.

Não contabilizaremos aqui a abundante informação enciclopédica contida na própria definição, uma vez que seria bastante difícil de quantificar²², salvo se esta estiver claramente delimitada (entre parênteses, por exemplo).

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

²¹ 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 [...]».

²² 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).

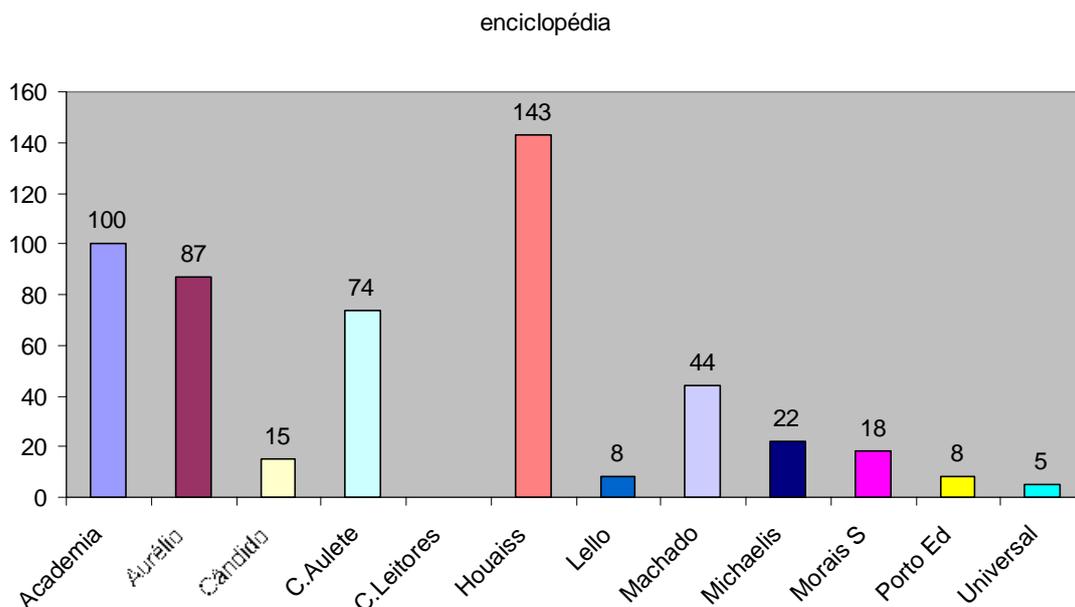


Gráfico 6
Etiquetas enciclopédico-cognitivas

4.7. Informação pragmático-contextual e retórica

Com "etiquetagem pragmático-contextual e retórica" referimo-nos a informações como:

- actos de fala: a função ou intenção retórico-comunicativa (pense-se por exemplo nas etiquetas do tipo *usado para*, 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: ARC., CAL., FAM., INF., POP.

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

- Informação sobre gestos que acompanham determinados enunciados ou enunciados que acompanham determinados gestos, como por exemplo a informação sobre o uso da fórmula de cortesia *Com licença*, empregue quando o falante vai rasgar um papel à frente do interlocutor (assim como ao desligar o telefone ou abrir um envelope).

- 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 figurado. 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 nossos dicionários:

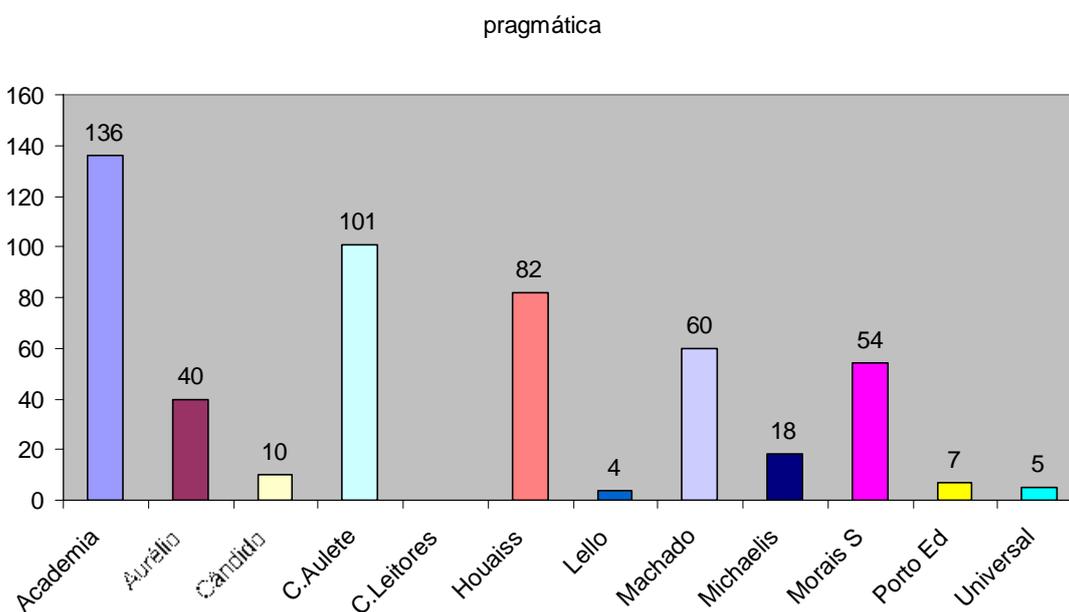


Gráfico 7

Etiquetas pragmático-contextuais e retóricas

5. Conclusões: " os melhores do teste "

Apresentaremos a seguir dois últimos gráficos com a soma, em forma de percentagem, dos valores de todos os quantitativos apresentados acima. Para estabelecer esta percentagem 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 percentagens²³.

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.

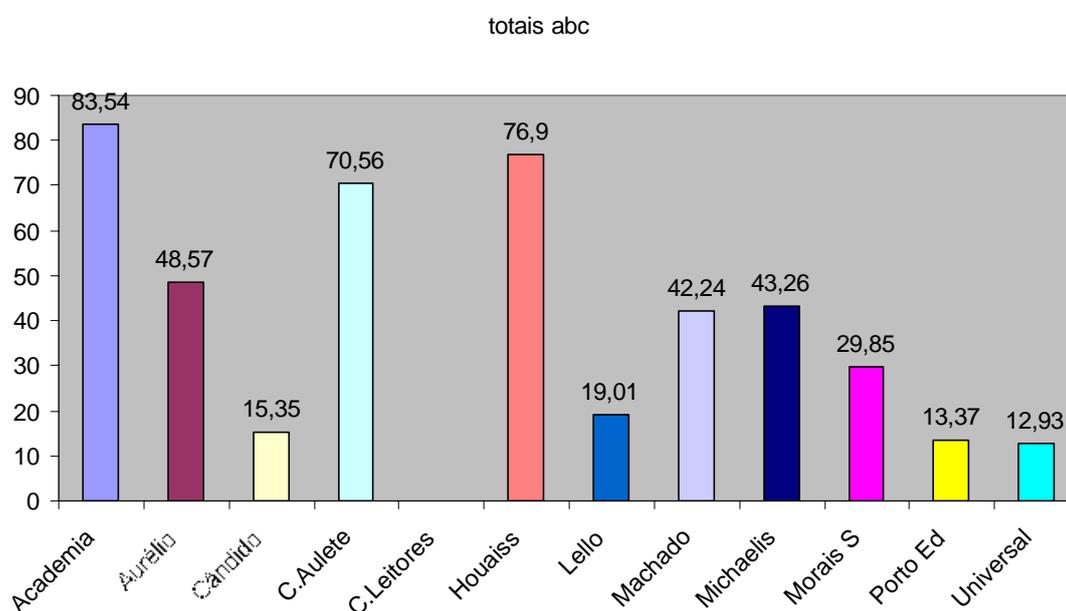


Gráfico 8

Totais

Reordenamos agora os resultados para que o leitor possa assim apreciar e comparar os resultados finais e descobrir quais são "os melhores dicionários do teste":

²³ Quero agradecer ao Engenheiro José João Almeida as suas sugestões e comentários para a elaboração do quadro com os valores totais final (vd. Anexo, quadro 9). Evidentemente, os eventuais erros ou faltas de exactidão são da minha inteira responsabilidade.

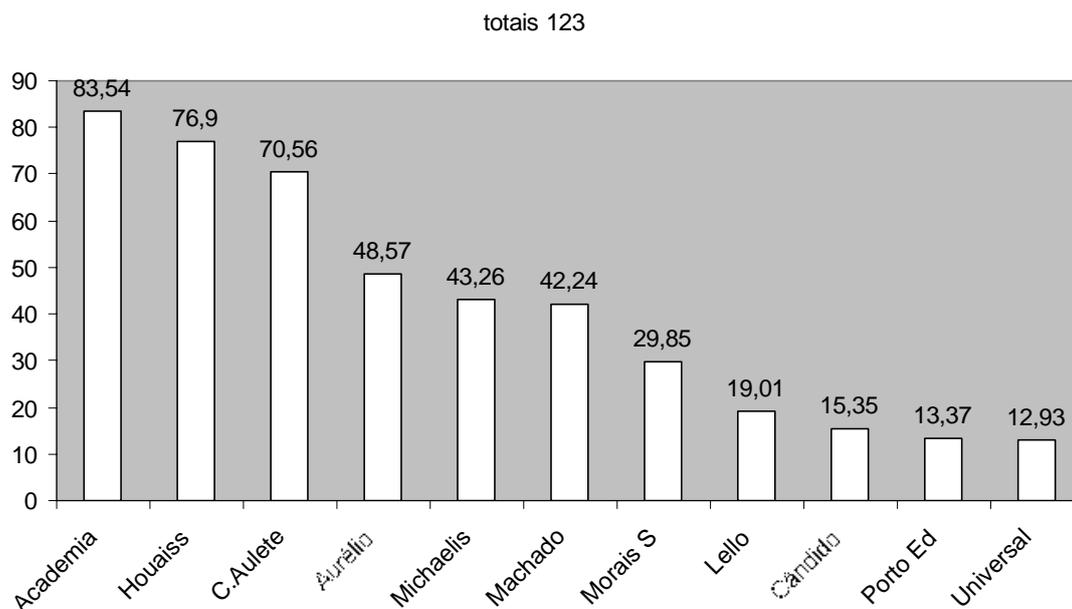


Gráfico 9

«O melhor do teste é...»

Talvez seja importante um comentário final relativo aos resultados obtidos pelo *Caldas Aulete*, que poderão surpreender a algumas pessoas. Este dicionário continua a ser um dos melhores dicionários portugueses no que se refere à combinatória lexical, fraseologia, exemplos e abonações²⁴. Mereceria por isso uma nova edição actualizada e muito mais cuidada no que se refere ao seu aspecto gráfico.

6. Referências Bibliográficas

6.1. Dicionários

ALMEIDA Costa J. e A. SAMPAIO E MELO

(2000): *Dicionário da Língua Portuguesa*. 8ª Edição, revista e ampliada. Porto: Porto Editora. [aqui: *Porto Editora*].

AULETE, F. J. Caldas

(1987): *Dicionário da Língua Portuguesa Caldas Aulete*. 5ª edição brasileira, revista, actualizada e aumentada por Hamílcar de Garcia e Antenor Nascentes. Rio de Janeiro: Editora Delta. [aqui: *Caldas Aulete*].

CASTELEIRO, J. Malaca (coord.)

²⁴ Assim já o indicava Rodrigues Lapa na sua *Estilística da Língua Portuguesa*:

«É precisamente neste capítulo da fraseologia, muito importante, que os dicionários correntes deixam mais a desejar. O mais celebrado de entre eles e o mais moderno dos grandes dicionários, o de Cândido de Figueiredo, é muito pobre em grupos fraseológicos, o que constitui um grave defeito, porque é nessas locuções que se imprime o chamado génio da língua. Como repositório de fraseologia, nada há que possa substituir entre nós o *Dicionário Contemporâneo* de Caldas Aulete.» (Lapa, 1984: 83).

(2001): *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*. Lisboa: Academia das Ciências de Lisboa/Editorial Verbo. [aqui: *Academia*].

FERREIRA, A. Buarque de Holanda

(1999): *Novo Aurélio Século XXI: O Dicionário da Língua Portuguesa*. Rio de Janeiro: Nova Fronteira. [aqui: *Aurélio*].

FIGUEIREDO, Cândido de

(1982): *Dicionário da Língua Portuguesa*. 16^a Edição. Lisboa: Livraria Bertrand. [aqui: *Cândido*].

HOUAISS A. (coord.)

(2001): *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objectiva. [aqui: *Houaiss*].

LELLO E IRMÃO EDITORES

(1996) *Novo Dicionário da Língua Portuguesa*. Sistema J. Amadora: Lello e Irmão Editores. [aqui *Lello*].

MACHADO, J. P.

(1981): *Grande Dicionário da Língua Portuguesa*. Lisboa: Sociedade de Língua Portuguesa / Amigos do Livro Editores. [aqui *Machado*].

SILVA, A. Morais da

(1980): *Novo Dicionário Compacto da Língua Portuguesa*. Lisboa: Editorial Confluência. [aqui *Morais*].

TEXTO EDITORA

(1995): *Dicionário Universal da Língua Portuguesa*. Lisboa: Texto Editora. [aqui: *Universal*].

VILELA, M.

(1991): *Dicionário do Português Básico*. Porto: Edições Asa. [aqui: *Vilela*]

WEISZFLOG, W.

(1998): *Moderno dicionário da Língua Portuguesa*. São Paulo: Melhoramentos. [aqui *Michaelis*].

6.2. *Bibliografia geral*

BESSE, H. & R. PORQUIER

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

CARVALHO, J. Herculano de

(1979): *Teoria da Linguagem. Natureza do Fenómeno Linguístico 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, L.

(1996): «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.

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.

LAPA, M. RODRIGUES

(1984): *Estilística da Língua Portuguesa*. Coimbra: Coimbra Editora.

MATHIEU-COLAS, M.

(1994): *Les mots à traits d'union. Problèmes de lexicographie informatique*. Paris: Didier.

ROCHA, P.A. & SANTOS, D.

(2000): «CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa», em *Actas do V Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR '2000)*, 12-19 de Novembro de 2000, págs. 131-140. Também em <http://www.lingateca.pt/Diana/download/RochaSantosPROPOR2000.pdf>

SOKAL, A & J. BRICMONT

(1999): *Imposturas Intelectuais*. Lisboa: Gradiva.

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.

WERNER, R.

(1982): «La unidad léxica y el lema», em Haensch *et al.* (1982), 188-232.