

Corpora textuais informatizados como ferramenta para análise da macroestrutura do Dicionário da Real Academia Espanhola

Computerized textual corpora as a tool to analyze the Dictionary of the Spanish Royal Academy's macrostructure

Manuela Arcos¹

DOI: 10.28998/2317-9945.2019n63p154-173

Resumo

O Dicionário da Real Academia Espanhola (DRAE) constitui um instrumento lexicográfico de referência na comunidade falante de língua espanhola. A Real Academia Espanhola (RAE) lança, no intervalo de dez anos, aproximadamente, uma nova edição de seu dicionário, com o propósito de oferecer uma imagem léxica atualizada do idioma espanhol. Para tanto, a RAE utiliza seu banco de dados, corpora textuais que registra o léxico diacrônico e sincrônico da língua. O objetivo deste trabalho é analisar quantitativa e qualitativamente, através dos corpora da Academia, a macroestrutura das duas edições mais recentes do seu dicionário (DRAE 2001; DRAE 2014).

Palavras-chave: Dicionário semasiológico. Dicionário monolíngue. Macroestrutura. Linguística de corpus

Abstract

The Diccionario de la Real Academia Española (DRAE) is a lexicographical instrument of reference for Spanish speaking community. The Real Academia Española (RAE) publishes about once every ten years a new edition of its dictionary with the purpose of giving an updated lexical image of the Spanish language. To that end, RAE uses its own data bank, that is, textual corpora that record the diachronic and synchronic lexicon of the language. The aim of this work is to analyze qualitatively and quantitatively, through the Academy's corpora, the macrostructure of the two most recent editions of the dictionary (DRAE 2001; DRAE 2014).

Keywords: Semasiological dictionary. Monolingual dictionary. Macrostructure. Corpus linguistics

Recebido em: 23/01/2019.

Aceito em: 30/03/2019.

¹ Mestranda em Linguística pelo Programa de Pós-graduação em Letras da Universidade Federal do Rio Grande do Sul. Graduada em Letras Português/Espanhol pela mesma instituição.

Introdução

A Real Academia Espanhola (RAE) é a instituição de autoridade no que se refere às questões linguísticas do idioma espanhol. Seu dicionário, o Dicionário da Real Academia Espanhola (DRAE), propõe-se a ser um dicionário geral da língua, e é uma obra de referência entre a comunidade falante. Com isso, o DRAE possui um valor normativo único em seu gênero, em virtude de sua importância social, de sua multiplicação editorial e de sua presença constante na comunidade, o que justifica o sentimento reverencial e de respeito que evoca em seus usuários (SALVADOR, 1992).

Em sua tarefa de oferecer um dicionário que funcione como um espelho da língua espanhola, a RAE oferece a cada dez anos, aproximadamente, uma nova edição do seu dicionário, de forma a oferecer uma imagem léxica fiel da língua. Contudo, observamos, mesmo nas edições recentes (atualizadas) do DRAE, a presença de palavras que nem sempre encontram um respaldo de registros que justifiquem sua lematização. Esse fator é causador de um aumento na densidade macroestrutural, que chamaremos “inchaço macroestrutural” (BUGUEÑO; FARIAS, 2008), que está relacionado às escolhas quantitativas e qualitativas tomadas para a constituição da macroestrutura de um dicionário.

Se compararmos a extensão macroestrutural do DRAE 2001 (22ª edição) e do DRAE 2014 (23ª edição), segundo a própria Academia, a 22ª edição conta com 88 mil entradas, aproximadamente, e a 23ª, com 93 mil, constando, assim, um aumento de 5 mil entradas de uma edição para a outra. Por outro lado, é a partir da 22ª edição do DRAE que a Academia passa a utilizar seu banco de dados (*corpora* diacrônicos e sincrônicos elaborados e oferecidos gratuitamente *on-line* pela instituição)². Nesse sentido, a tarefa de atualização dos dicionários passa a contar com um grande contingente de textos informatizados (*corpora* textuais) cujo objetivo é oferecer uma amostra da língua em sua realização efetiva. As informações oferecidas por esses *corpora* permitem o acesso a informações diacrônicas, diatópicas e diamésicas dos textos, além de outros dados, como a frequência de aparição de uma forma lexical e seus padrões colocacionais. Entre esses eixos, usaremos para nossa análise o eixo diacrônico, ou seja, observaremos a marcação temporal das entradas do dicionário nos *corpora* da Academia.

Contudo, há que destacar que, no que se refere à macroestrutura, ainda não há, na literatura especializada, um consenso quanto ao seu tamanho em um dicionário geral de língua. Segundo Hasmann e Wiegand (2003, p. 337), “há pouquíssimos estudos empíricos sobre perfis macroestruturais de dicionários gerais monolíngues, e também pouca descrição operacional que permita medir a riqueza da macroestrutura”³ (tradução nossa)³.

Com isso, o objetivo deste trabalho é, partindo de uma breve revisão teórica sobre macroestrutura, propor uma metodologia de uma análise quantitativa e qualitativa deste componente canônico do dicionário que parta de dados empíricos, como são os oferecidos pelos *corpora* textuais. Julgamos que o estudo seja válido por contribuir para a discussão sobre a extensão macroestrutural do dicionário geral da língua, uma vez constatada a lacuna de estudos deste tipo no âmbito da Lexicografia.

² Informações apresentadas no *frontmatter* do DRAE (2001) e do DRAE (2014).

³ No original: “There are very few empirical studies on macrostructure profiles of general monolingual dictionaries and also little operational description to enable us to measure the richness of macrostructure” (HAUSMANN; WIEGAND, 2003, p. 337).

O dicionário geral de língua

Segundo Biderman (1998), o dicionário se arquiteta como uma organização sistemática do léxico e como um depositário do acervo lexical da cultura. Desse modo, o dicionário é a memória lexical de uma sociedade, posto que “constitui o acervo e o registro das significações que a nossa memória não é capaz de memorizar” (BIDERMAN, 2004, p. 185). Para a autora, os dicionários monolíngues se apresentam em diversos tipos: os dicionários de língua, os dicionários ideológicos, os dicionários temáticos ou especializados, os dicionários etimológicos, os dicionários históricos, os dicionários terminológicos, entre outros. Sobre os dicionários de língua, ou dicionário “geral” da língua, em especial, a sua função é recolher o vocabulário de uma língua num determinado momento da história de um grupo social (BIDERMAN, 1998, p. 130).

Para Bugueño (2013), há três parâmetros que se relacionam diretamente com as estruturas informativas do dicionário: a classe do dicionário, a função que irá cumprir e o perfil de usuário. A partir desses três parâmetros, Bugueño (2014, p. 228) propõe uma taxonomia de obras lexicográficas na qual classifica o dicionário geral como um dicionário semasiológico, monolíngue, cuja função é oferecer uma imagem léxica exaustiva e diassistemicamente inclusiva⁴ da língua, e cujo usuário é o falante nativo.

Classificar um dicionário como semasiológico significa dizer que o dicionário organiza suas informações partindo do lema para a definição, ou, numa visão saussuriana do signo, do significante para o significado. A direção significante → significado acarreta que a função do dicionário semasiológico seja de compreensão, uma vez que o usuário parte do significante, isto é, já conhece a forma concreta, e almeja chegar ao seu significado, ou seja, sua designação. Além da direção significante/significado, outra característica do dicionário semasiológico é a sua ordenação dos lemas⁵. Segundo Svensén (2009, p. 22), no dicionário semasiológico, a organização alfabética representa o tipo mais importante de ordenação macroestrutural.

Em sua taxonomia, Bugueño (2013) propõe três parâmetros – classe, função e usuário – que serão decisivos para a configuração dos componentes canônicos do dicionário, isto é, os que estarão presentes obrigatoriamente na obra lexicográfica. O autor estabelece que para o dicionário geral de orientação semasiológica existem quatro componentes canônicos: a macroestrutura, a microestrutura, a medioestrutura e o *front matter* (2013, p. 22). A macroestrutura, também chamada de “nomenclatura” do dicionário por Biderman (1998), é a progressão vertical dos lemas, e se refere ao número de verbetes que o dicionário contém. A microestrutura é o conjunto de informações sobre o signo- lema, e possui uma função predicativa em relação ao lema (BUGUEÑO; FARIAS, 2006 *apud* BUGUEÑO, 2013, p. 22). A medioestrutura consiste no sistema de remissões, cujo propósito é direcionar o usuário rápida e efetivamente à informação que o dicionário pretende fornecer. Por fim, o *front matter* deve informar a qual usuário o dicionário se destina, a função do dicionário, e também servir como um manual de instruções e de informações. Uma vez que o propósito deste trabalho é traçar uma análise da

⁴ Isto é, o dicionário semasiológico inclui o máximo de eixos do diassistema da língua (COSERIU, 1980), como o eixo diacrônico (temporal), diatópico (geográfico), diafásico (de uso) e diastrático (estrato social). No que se refere ao eixo diacrônico, como citado anteriormente de acordo com Biderman (1998), o dicionário geral recolhe o vocabulário de um determinado momento da língua; portanto, o eixo diacrônico não é exaustivo, mas “α exaustivo”, na taxonomia de Bugueño (2014).

⁵ Ver também conceito de “estrutura de acesso” em Hausmann e Wiegand (2003), Wiegand e Beer (2013) e Bugueño (2004/2005).

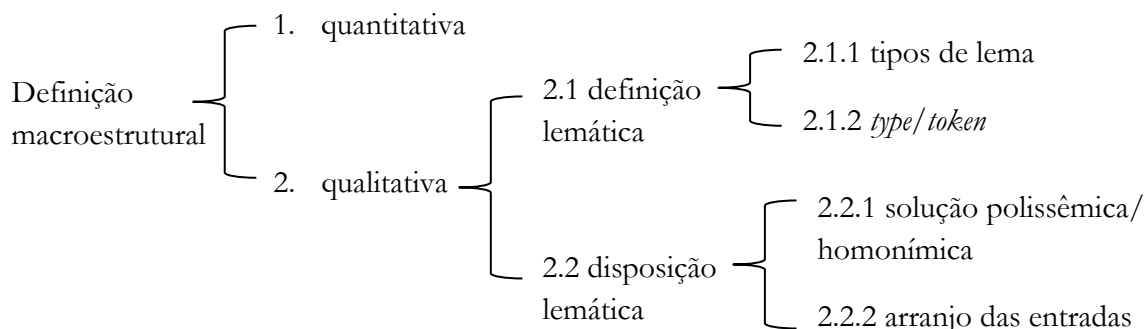
macroestrutura do DRAE, na seção seguinte aprofundamos os conceitos referentes a esse componente.

Componentes canônicos do dicionário: a macroestrutura

De acordo com Bugueño e Farias (2008, p. 137-138), a macroestrutura é o “estabelecimento do número de verbetes que o dicionário conterà, assim como o número de unidades passíveis de lematização”. No que se refere aos dicionários gerais, Hausmann e Wiegand (2003) explicam que a macroestrutura é o componente do dicionário que determina sob qual lema um item lexicográfico será encontrado. Quanto à organização macroestrutural, os mesmos autores (2003, p. 336) indicam que a característica mais comum no desenho da macroestrutura é a orientação de cima para baixo (*from top to bottom*), com uma ordenação alfabética em que o lema é arranjado verticalmente.

A macroestrutura do dicionário pode ser vista sob dois ângulos: quantitativo e qualitativo. Bugueño e Farias (2008) oferecem uma visão aprofundada da macroestrutura nesses dois aspectos, conforme o esquema ilustrado abaixo, proposto pelos autores.

Esquema 1: definição macroestrutural.



Fonte: Bugueño e Farias (2008, p. 138).

A partir desse modelo, desenvolvemos e comentamos brevemente cada aspecto das definições macroestruturais quantitativa e qualitativa nas seções que seguem.

Definição macroestrutural quantitativa

A definição macroestrutural quantitativa corresponde ao cálculo de lemas a serem arrolados no dicionário que, por sua vez, corresponde ao que os autores chamam de “densidade macroestrutural”. Esse cálculo é válido para que se evite o inchaço macroestrutural, isto é, a lematização de uma grande quantidade de lemas que possuem pouca ou nenhuma utilidade para o usuário. Para Bugueño (2014), a discussão sobre o cálculo macroestrutural fundamentado matematicamente para cada tipo de dicionário ainda se encontra em seus primórdios. Contudo, para o autor, o inchaço macroestrutural se deve a uma apresentação numericamente aleatória de uma grande quantidade de lemas, isto é,

não há um critério ou cálculo que justifique o número de lemas que são incluídos na nomenclatura do dicionário geral de língua.

Ainda sobre o assunto, Farias (2010) constata que o inchaço macroestrutural está relacionado à inclusão em massa de palavras diatopicamente marcadas, à lematização de unidades desusadas e pouco frequentes, além da inclusão de termos técnicos, isto é, não pertencentes à língua geral. Embora a autora (FARIAS, 2010, p. 83) defenda que um dicionário para a compreensão – como é o caso do DRAE – deva ser macroestruturalmente denso, há determinadas unidades léxicas que poderiam ser dispensadas.

Uma forma de evitar o inchaço da nomenclatura do dicionário de língua é estabelecer critérios que determinem quais unidades léxicas serão registradas. Bugueño (2004/2005, p. 21) propõe dois critérios para a seleção macroestrutural: o estatístico e o(s) critério(s) sin- ou diassistêmicos. No caso do critério estatístico, Bugueño apresenta como exemplo o dicionário *Collins Cobuild* (1995), cuja tradição consiste em ancorar sua seleção macroestrutural numa representação estatística do léxico. Desse modo, uma escala de frequência de unidades é estabelecida, calculando-se um número mínimo de ocorrências. Assim, uma unidade com frequência menor que o número mínimo de ocorrências estabelecido previamente não será lematizada. Quanto aos critérios sin- ou diassistêmicos, o autor explica que se pode utilizar a sincronia como variável para determinar a densidade léxica da macroestrutura, sendo os intervalos de tempo mais empregados os de algumas dezenas de anos – como é o caso do DRAE. Além do parâmetro sincrônico, outros podem ser empregados, como optar por uma perspectiva sintópica frente a uma diatópica, uma sinstrática por uma diastrática etc. (BUGUEÑO, 2004/2005, p. 22).

Definição macroestrutural qualitativa

A definição macroestrutural qualitativa se divide em definição lemática (2.1) e disposição lemática (2.2). A primeira permite determinar quais tipos de unidades podem fazer parte da nomenclatura de um dicionário, enquanto a segunda se refere à disposição dessas unidades, isto é, de que forma serão apresentadas (BUGUEÑO; FARIAS, 2008).

Na definição lemática decide-se, por exemplo, se o dicionário lematizará unidades lexicais simples, unidades fraseológicas, formas presas (prefixos e sufixos), siglas e abreviaturas. Ou, ainda, se o dicionário lematizará uma unidade pela sua forma *type* (forma canônica, de maior prestígio) ou *token* (variante, de menor prestígio) (BUGUEÑO; FARIAS, 2008, p. 152).

Já na disposição lemática, decide-se se o dicionário optará por uma solução polissêmica (lematização de formas homônimas homógrafas ou homófonas sob um único verbete) ou por uma solução homonímica (lematização das formas homônimas em verbetes separados) (FARIAS, 2009, p. 106). Para Landau (2001, p. 100 *apud* FARIAS, 2009, p. 106), o tipo de solução escolhida, polissêmica ou homonímica, afeta a macroestrutura, uma vez que interfere diretamente na contagem das entradas.

Outra escolha na disposição lemática será o tipo de arranjo das entradas, como a ordenação alfabética. Wiegand (1989 *apud* BUGUEÑO, 2004/2005) classifica a disposição lemática em três possíveis formas: estrutura lisa, estrutura de nicho léxico e estrutura de

ninho léxico. A estrutura lisa é a adotada pela maioria dos dicionários, na qual as unidades estão lematizadas em progressão vertical de ordenação alfabética e recuadas à esquerda. Na estrutura de nicho léxico se lematizam de maneira integrada num único bloco todas as unidades que possuem uma base etimológica comum a uma unidade léxica primitiva (*simplex*), que constituirão sub-entradas alfabeticamente ordenadas e com definição. Por fim, na estrutura de ninho léxico se lematizam unidades também de forma subordinada ao *simplex*, mas, sem respeito à ordem alfabética (de forma regressiva em relação à ordem alfabética, conforme Bugueño (2004/2005, p. 27)) e sem definição.

Para levar a cabo nossa análise e observar como se constitui a macroestrutura no DRAE, levaremos em conta a análise quantitativa da macroestrutura a partir da metodologia que proporemos. No que concerne à análise qualitativa, analisaremos somente a disposição lematizada (2.2.1 e 2.2.2 do Esquema 1), por considerarmos esse parâmetro como um possível responsável por um inchaço da nomenclatura no DRAE. Antes de introduzir nossa proposta metodológica de análise, apresentamos, na seção seguinte, o DRAE e os *corpora* da Real Academia Espanhola.

O DRAE e os *corpora* da RAE

Para Grijelmo (2001, p. 368), a Real Academia Espanhola (RAE) ainda desfruta de um sólido prestígio em todo o mundo hispano. A seu dicionário (DRAE) se dirigem os falantes que se deparam com uma dúvida ou com uma curiosidade, fato que confere à Academia um papel de autoridade institucional em tudo aquilo relativo ao idioma espanhol. Desse modo, a comunidade falante de língua espanhola conta com a RAE como uma entidade que orienta todas as decisões e questões linguísticas do idioma. O lema da Academia – refletido no papel de seu dicionário – desde seu surgimento, em 1713, é *limpia, fija y da esplendor*. Nesse sentido, “limpar” significa manter um equilíbrio entre o léxico vigente, o léxico emergente (neologismos ou estrangeirismos) e o léxico em desuso. “Fixar” significa consagrar unidades que encontrem um uso real na comunidade. Finalmente, “dar esplendor” significa, atualmente, manter a memória na comunidade falante da historicidade da língua.

O dicionário da Academia é, portanto, responsável por executar os propósitos da instituição: limpar, fixar e dar esplendor. Para Salvador (1992), seu valor normativo reside na sua importância social como dicionário, na sua multiplicação editorial e na sua presença constante entre a comunidade linguística, o que justifica o sentimento reverencial e respeitoso que evoca em seus usuários. Por isso, o DRAE terá sempre “a última palavra sobre todas as palavras” (SALVADOR, 1992, tradução nossa).

Na sua tarefa de oferecer uma imagem léxica atualizada da língua espanhola, a Academia lança novas edições de seu dicionário a cada dez anos, aproximadamente. Nesse sentido, desde a 22ª edição (DRAE, 2001), a RAE já anunciava o uso de seu Banco de Dados para fins de atualização do dicionário. Assim, encontramos no *front matter* do DRAE 2001 a seguinte nota:

La revisión afecta también, como es lógico, a lo ya incluido, y se completa con numerosas adiciones. El trabajo desarrollado en este punto durante los diez años transcurridos ha sido, en ese doble aspecto, intensísimo. Las cifras hablan por sí solas: dos tercios de los artículos registrados en la anterior edición han sido enmendados en mayor o

menor medida, 55.442 exactamente, y a ellos se han añadido 11.425 nuevas entradas, 24.819 nuevas acepciones y 3.896 formas complejas. (...) Toda esta tarea ha sido posible, en buena medida, gracias al *Banco de datos del español* que la Academia ha construido durante este último decenio y que en estos momentos, en su doble repertorio, histórico y actual – *Corpus diacrónico del español* (CORDE) y *Corpus de referencia del español actual* (CREA)⁶ – supera los 270 millones de registros léxicos (DRAE, 2001, p. X).

Com isso, desde a edição de 2001, a RAE já contava com recursos informáticos para a tarefa de atualização de seu dicionário quanto à inclusão de novas entradas e exclusão de outras, à reformulação de aceções e à marcação diassistêmica do léxico. Atualmente, a Academia conta com mais um *corpus* além do CORDE e do CREA, o CORPES XXI (*Corpus del Español del Siglo XXI*), lançado em 2013. Enquanto o CORDE compreende o léxico da língua desde seus primeiros registros até o ano de 1974, o CREA registra o léxico no intervalo de 1975 até 2004. Já o CORPES XXI, surge como um complemento ao CREA, e atualmente conta com 225 milhões de formas de registros, que vão de 1975 a 2016. Entre as melhorias do CORPES XXI em relação ao CREA estão a abrangência cronológica (o CREA oferece registros até o ano de 2004), a abrangência diatópica (70% dos registros oriundos da América, 30% da Espanha, enquanto no CREA a divisão é 50%/50%, oferecendo dados distorcidos da realidade linguística do idioma) e, por fim, o fato de ser um *corpus* anotado, que permite que se façam buscas de formas, lemas (formas canônicas) e de unidades complexas (sintagmas, expressões, etc.).

Sobre o DRAE 2014, uma vez que analisamos a versão *on-line*, na sessão correspondente ao *frontmatter* (RAE, 2019) a Academia apresenta informações da forma como o dicionário é atualizado, descrevendo as fontes que utiliza para tal tarefa:

La preparación de cada nueva edición del diccionario académico implica, por tanto, la identificación de nuevas palabras o nuevos significados y la revisión de las palabras que ya figuraban en él. Esta revisión se realiza agrupando las palabras según diversos criterios, como el área temática a la que pertenecen (palabras de la música, palabras de la química...), por sus características gramaticales (sustantivos, conjunciones, adjetivos invariables...), por su procedencia (extranjerismos, revisión de etimología...), por el área geográfica en la que se documentan, etc. (...) Las fuentes documentales del diccionario académico, que se han ido creando y ampliando en distintas etapas de su historia, son actualmente las siguientes: El banco de datos del español, que cuenta con más de 400 millones de registros de textos históricos y actuales de todos los países hispanohablantes. El fichero histórico de la Academia, con sus más de diez millones de papeletas léxicas y lexicográficas. Las obras de referencia y estudios monográficos sobre léxico. La Unidad Interactiva (UNIDRAE), un servicio creado para atender y canalizar las propuestas y sugerencias externas relacionadas con el *Diccionario*. Gracias a estas fuentes y a las mejoras desarrolladas en su forma de consulta y utilización, los lexicógrafos pueden hoy por hoy obtener los datos precisos para la toma de decisiones de carácter normativo y para la preparación de las propuestas que posteriormente estudiarán las comisiones académicas (RAE, 2019, *on-line*).

⁶ Neste trabalho, não usaremos o CREA como banco de dados para análise da macroestrutura, mas o CORPES XXI (*Corpus del Español del Siglo XXI*), novo banco de dados, mais amplo e com recursos mais modernos, lançado pela RAE em 2013.

Desse modo, a Academia afirma a importância da sua base de dados no processo de elaboração e atualização das novas edições do seu dicionário, e a apresenta como uma ferramenta que faz parte dessa tarefa.

Na seção seguinte, explicamos a metodologia que propomos para analisar a macroestrutura do DRAE, comparando a 22^a e 23^a edições com base nos *corpora* da própria Academia.

Metodologia

Como dito anteriormente, nosso foco neste trabalho é analisar a pertinência das unidades lematizadas na macroestrutura do DRAE. Para isso, comparamos as duas últimas edições do dicionário, a 22^a edição (DRAE, 2001) e a 23^a edição (DRAE, 2014). Optamos por comparar as duas edições para observar de que forma a atualização da macroestrutura do dicionário é feita pela RAE. Desse modo, seguimos os seguintes passos:

- 1) Selecionamos aleatoriamente dois intervalos lematizados de 50 palavras do DRAE 2001. O primeiro intervalo foi da entrada *aarónico* até *abastecer*, e o segundo intervalo foi da entrada *hazmerreír* até *bediento*⁷.
- 2) Os mesmos intervalos foram recolhidos no DRAE 2014, embora o número de verbetes tenha variado em função da inclusão e exclusão de alguns verbetes, como comentaremos na análise.
- 3) Pesquisamos cada lema no CORDE e no CORPES XXI, a fim de observar a frequência de aparição, isto é, o número de ocorrências da palavra tanto na língua atual (CORPES XXI) quanto em estágios anteriores do idioma (CORDE)⁸.
- 4) Observamos, para ambos os intervalos, a pertinência das marcações cronológicas (*antiquada* e *desusada*).

Para analisar a marcação cronológica, tomamos como referência o critério da própria Academia, apresentado no *frontmatter* de ambas as edições:

- Lemas marcados como **antiquados** (*ant.*) são palavras que pertencem à Idade Média (até o século XV), mas que foram descartadas na língua moderna.
- Lemas marcados como **desusados** (*desus.*) são palavras usadas na Idade Moderna (século XV a XVIII) que hoje já não se empregam mais.
- Lemas sem marcação são palavras que pertencem ao período da língua que o DRAE se propõe registrar (intervalo do século XVIII até a atualidade).

⁷ Todos os intervalos analisados das duas edições do DRAE se encontram no anexo ao final deste trabalho.

⁸ O CREA não foi incluído na análise, uma vez que as marcas *antiguado* e *desusado* se encontram no período registrado apenas no CORDE. A comparação com os dados do CORPES XXI serve para confirmar se essas palavras ainda constam no espanhol atual, podendo, dessa forma, estar com uma marcação cronológica inadequada. Outro motivo para não analisarmos os dados do CREA se deve ao fato do CORPES XXI abranger todos os dados que o CREA apresenta, uma vez que o período dos registros do primeiro vão de 1975 a 2016 e, do segundo, de 1975 a 2004.

Ademais, observamos se há palavras que constituam *ghost words*⁹, isto é, formas que não apresentam ocorrência nenhuma no *corpus*, e também se há presença de *hápax legomena*¹⁰, isto é, palavras que apresentam apenas uma ocorrência.

A Tabela 1 ilustra como o levantamento dos dados foi feito: partimos dos verbetes do intervalo lematizado selecionado e observamos se estão presentes na macroestrutura somente da edição de 2001 ou 2014, ou, ainda, em ambas. Em seguida, pesquisamos as palavras nos dois *corpora* da Academia, CORDE e CORPES XXI, e registramos sua ocorrência. Quando constatamos erros na marcação cronológica, registramos na coluna “CORDE” o ano dos registros em que a palavra aparece registrada. Por fim, na coluna “análise”, colocamos as observações que anotamos acerca do verbo, por exemplo, se é uma *ghost word*, uma *hápax legomena*, se a marcação cronológica está em desacordo com os critérios da Academia ou se está adequada.

Tabela 1 – Análise dos verbetes.

Verbo	DRAE 2001	DRAE 2014	CORPES XXI	CORDE	Análise
Ababillarse	✓	✓	0	0	<i>Ghost word</i>
Abarrenar (<i>desus.</i>)	✓	✗	0	1 (1582)	<i>Hápax legomena</i> ;
Abarrado,da	✓	✓	0	2 (1575/1646)	Marca <i>desus.</i>
Abad	✓	✓	695	5088	Marcação cronológica adequada

Fonte: autora deste artigo.

Julgamos importante o uso do CORDE para a atualização do dicionário, uma vez que, por mais que sejam palavras desusadas ou antiquadas, se apresentarem ocorrências, justifica-se sua lematização. Assim, se o usuário se depara com essas palavras, o dicionário oferece a possibilidade de consulta. Contudo, registrar palavras que não apresentam ocorrências nem ao menos no CORDE (mesmo que estejam marcadas como antiquadas ou desusadas) pode corroborar para o inchaço da macroestrutura do dicionário, como aprofundaremos mais adiante.

Além da análise quantitativa da macroestrutura descrita anteriormente, fazemos uma breve observação qualitativa no que se refere ao tipo de disposição lematizada que o DRAE segue (solução homônima ou polissêmica e tipo de arranjo macroestrutural).

⁹ Para Larson (2010), há diferentes motivos para que uma palavra se apresente como uma *ghost word*. Entre eles, os mais frequentes são erros próprios do fazer lexicográfico (como erros de digitação ou a intuição do lexicógrafo) que, com a reprodução editorial, acabam se perpetuando nos dicionários. Outra possibilidade que o autor considera para as *ghost words* é a busca de uma forma num *corpus* com registros que não condizem cronologicamente com o momento de uso da palavra, como, por exemplo, pesquisar uma palavra em desuso num *corpus* sincrônico. Por isso, na nossa metodologia optamos por consultar tanto o *corpus* diacrônico quanto o sincrônico oferecidos pela RAE, com o objetivo de não julgarmos erroneamente que uma forma se trata de uma *ghost word*.

¹⁰ Uma vez que os *corpora* utilizados neste trabalho são complementares, isto é, os registros que constituem o CORDE não são os mesmos que constituem o CORPES XXI, consideramos *hápax legomena* neste trabalho as palavras que apresentarem apenas um registro em somente um dos *corpora*, seja ele o diacrônico ou o sincrônico. Para mais informações sobre o conceito de *hápax legomena*, ver Malmkjær (2004).

Na próxima seção, ilustramos a análise dos dados e a quais resultados quantitativos e qualitativos chegamos.

Análise e resultados da macroestrutura do DRAE

a) Análise quantitativa

Como já explicado na metodologia, iniciamos a análise partindo de dois intervalos lematícos de 50 lemas cada do DRAE 2001, das palavras *aarónico* até *abastecer* e de *hazmerreír* até *bediento,ta*. O propósito de partir da edição de 2001 foi observar se, ao comparar com o mesmo intervalo na edição de 2014, ocorreriam mudanças na macroestrutura nos seguintes aspectos:

- Aumento ou diminuição da macroestrutura pela inclusão ou exclusão de entradas;
 - *Ghost words* e *hápax legomena* lematizadas;
 - Entradas com problema na marcação cronológica.
- A seguir, descrevemos a análise e os resultados de cada um desses aspectos.

Inclusões e exclusões de entradas na macroestrutura

Ao comparar o intervalo lematíco de *aarónico* até *abastecer* do DRAE 2001 e do DRAE 2014, observamos que houve uma diminuição numérica de uma entrada na macroestrutura – o mesmo intervalo na edição de 2014 apresentou 49 lemas. Ao compará-los, constatamos que 11 entradas foram removidas e 10 novas acrescentadas na edição 2014.

Analisando as entradas que foram retiradas, constatamos que cinco se tratava de *ghost words*, a exemplo de *abachar* e *abadernar*, isto é, não apresentavam nenhuma ocorrência nos *corpora* e, portanto, coerentemente, não foram incluídas na nova edição. Por outro lado, das 10 novas entradas que a Academia incluiu no DRAE 2014, duas eram *ghost words*: *abadernar* e *abarquillamiento*.

Com relação ao intervalo de 50 entradas de *hazmerreír* até *bediento* da edição de 2001, no DRAE 2014 esse mesmo intervalo apresentou um crescimento macroestrutural, registrando 54 entradas. Desse modo, nenhuma entrada foi retirada e quatro foram acrescentadas: *heavy*, *heavy metal*, *hechizante* e *hectopascal*. É importante ressaltar que, apesar da Academia não ter retirado nenhuma entrada desse intervalo – embora apresentasse cinco *ghost words* e três *hápax legomena*¹¹ –, as entradas que foram incluídas parecem estar fundamentadas numa aparição considerável nos seus bancos de dados: *heavy* apresentou 179 ocorrências no CORPES XXI; *heavy metal*, 198; *hechizante*, 34; e *hectopascal*, 19. Desse modo, o aumento da macroestrutura se justifica pela ocorrência desses lemas com ocorrência de uso, embora, por outro lado, não se justifique pela presença das cinco formas sem nenhuma ocorrência.

¹¹ As cinco *ghost words* são *hebén*, *hebetar*, *bebijón*, *beciento*, *bectiquez*. As três *hápax legomena* são *hebdomadariamente*, com um registro do ano 1923; *hebroso*, com um registro de 1678 e, *hachiveresco*, com um registro de 1615.

Levando em conta todos os dados, observamos que os dois intervalos que analisamos do DRAE 2001 – totalizando 100 entradas – sofreram um aumento macroestrutural de três entradas para o DRAE 2014 (104 entradas), considerando as remoções e inclusões de novos verbetes, constituindo, pois, um aumento macroestrutural de 4% nos mesmos intervalos de uma edição para a outra. Vale observar que, embora a lematização de novas entradas no DRAE 2014 pareça estar respaldada pelos dados dos *corpora* da própria Academia (como no caso de *heavy*, *heavy metal*, *hechizante* e *hectopascal*), também observamos, por exemplo, a inclusão de duas entradas que constituem *ghost words* (*abadernar* e *abarquillamiento*), casos que serão comentados de maneira mais aprofundada a seguir.

Com isso, esses dados sugerem que a Academia nem sempre se baseia em seu banco de dados para a atualização de seu dicionário, sem constatar, portanto, se são palavras que ocorrem na língua espanhola do período que seus *corpora* abrangem.

Ghost words e hápax legomena lematizadas

Com relação aos resultados numéricos, das 100 entradas analisadas do DRAE 2001 constatamos 15 *ghost words* e nove *hápax legomena*. No DRAE 2014, entre as 104 entradas, identificamos 12 *ghost words* e sete *hápax legomena*. A Tabela 2 ilustra alguns exemplos desses resultados¹².

Tabela 2 – *Ghost words e hápax legomena lematizadas*.

Verbete	DRAE 2001	DRAE 2014	CORPES XXI	CORDE	Análise
Abácora (<i>ant</i>)	✓	✓	0	0	<i>Ghost word</i>
Abarquillamiento	✗	✓	0	0	<i>Ghost word</i>
Hechiceresco, ca	✓	✓	0	1	<i>Hápax legomena</i> ; Somente um (01) registro na forma feminina.
Hebdomariamente	✓	✓	0	1	<i>Hápax Legomena</i>

Fonte: autora deste artigo.

Entre os casos de *hápax legomena*, identificamos também dois casos cujo registro apareceu somente no gênero feminino: *hechiceresca* e *bedentinosa*. Contudo, os verbetes não apresentavam nenhum comentário que indicasse seu uso no feminino. Além disso, no caso de *hechiceresco,ca* o único registro que encontramos no CORDE constava do ano de 1615. Desse modo, o lema deveria estar marcado como *desusado*, conforme comentamos na seção seguinte.

Como resultado quantitativo final, identificamos 15 *ghost words* nos 100 verbetes analisados do DRAE 2001 e 12 *ghost words* nos 104 verbetes analisados do DRAE 2014. Desse modo, observamos um inchaço macroestrutural no que se refere à inclusão de *ghost words*, resultando num aumento de 15% no DRAE 2001 e de, aproximadamente, 12% no

¹² Os dados completos se encontram no anexo deste artigo.

DRAE 2014, constituindo uma diminuição de 3% da edição mais antiga para a mais recente. Esquematizamos os resultados na Tabela 3:

Tabela 3 – *Ghost words* no DRAE 2001 e 2014.

<i>Ghost words</i> no DRAE 2001		<i>Ghost words</i> no DRAE 2014	
Intervalo lematítico	Intervalo lematítico	Intervalo lematítico	Intervalo lematítico
<i>aarónico – abastecer</i>	<i>hazmerreír – bediento,ta</i>	<i>aarónico – abastecer</i>	<i>hazmerreír – bediento,ta</i>
10 de 50	5 de 50	7 de 49 ¹³	5 de 54
Total = 15 de 100 (15%)		Total = 12 de 104 (11,6%)	

Fonte: autora deste artigo.

Com relação às *hápax legomena*, isto é, verbetes que apresentam somente uma ocorrência nos *corpora*, identificamos um total de nove em 100 entradas do DRAE 2001 e sete em 104 entradas do DRAE 2014; portanto três *hápax* foram retiradas de uma edição para a outra. A Tabela 4 ilustra os dados numéricos.

Tabela 4 – *Hápax legomena* no DRAE 2001 e 2014.

<i>Hápax legomena</i> no DRAE 2001		<i>Hápax legomena</i> no DRAE 2014	
Intervalo lematítico	Intervalo lematítico	Intervalo lematítico	Intervalo lematítico
<i>aarónico – abastecer</i>	<i>hazmerreír – bediento,ta</i>	<i>aarónico – abastecer</i>	<i>hazmerreír – bediento,ta</i>
5 de 50	4 de 50	3 de 49	4 de 54
Total = 9 de 100 (9%)		Total = 7 de 104 (6,7%)	

Fonte: autora deste artigo.

Os dados numéricos sugerem uma reflexão interessante acerca da presença desse tipo de palavras na macroestrutura do DRAE. Observamos que, nos dois intervalos para ambas as edições, as proporções percentuais são muito semelhantes, inclusive na diminuição de *ghost words* e *hápax legomena* de uma edição para a outra – 15% e 12% (aprox.) de *ghost words* e 9% e 7% (aprox.) de *hápax legomena* nas edições 2001 e 2014, respectivamente.

Desse modo, considerando os dois intervalos e o número total de verbetes de cada edição, poderíamos presumir que o DRAE 2001 (88 mil entradas) apresenta um inchaço macroestrutural causado pelas *ghost words* de aproximadamente 13.200 entradas, e o DRAE 2014 (93 mil entradas) um total de 11.200 entradas, aproximadamente. Fazendo o mesmo cálculo proporcional para as *hápax legomena*, teríamos um aumento da macroestrutura de aproximadamente 8 mil entradas no DRAE 2001 e de 6,5 mil entradas no DRAE 2014.

Se considerarmos as *ghost words* e as *hápax legomena* conjuntamente, obteríamos, aproximadamente, 21.200 entradas no DRAE 2001 e 17.700 entradas no DRAE 2014 cuja lematização deveria ser revista pela Academia.

¹³ Ressaltando que os intervalos lematíticos do DRAE (2014) apresentam 49 e 54 entradas, posto que consideramos o mesmo intervalo escolhido para o DRAE (2001).

Entradas com problemas na marcação cronológica

Nossa análise permitiu constatar diferentes problemas no aspecto da marcação cronológica das entradas. A partir dos intervalos lematícos, identificamos três tipos de resultados: palavras que deveriam estar marcadas cronologicamente como *antiguadas* ou *desusadas*, palavras que não deveriam carregar marca cronológica, uma vez que os dados dos *corpora* indicam que são formas do léxico atual, e palavras corretamente marcadas. A Tabela 5 ilustra exemplos desses resultados.

Tabela 5 – Análise da marcação cronológica.

Entrada	Marcação cronológica	Data dos registros	Análise
Abarrajar	x	CORDE – 1 registro de 1535	Deveria levar marca de <i>desusada</i>
Abastanza	<i>desusada</i>	CORDE – 14 registros de 1541 a 1948 CORPES XXI – 1 registro de 2009	Não deveria levar a marca de <i>desusada</i>
Abarredera	<i>antiguada</i>	CORDE – 2 registros de 1251 e 1330	Marcação cronológica adequada

Fonte: autora deste artigo.

A partir dessa análise, identificamos 12 entradas com problemas na marcação cronológica no DRAE 2001 e 13 no DRAE 2014, conforme ilustra a Tabela 6.

Tabela 6 – Marcação cronológica no DRAE 2001 e 2014.

Marcação cronológica no DRAE 2001		Marcação cronológica no DRAE 2014	
Intervalo lematíco	Intervalo lematíco	Intervalo lematíco	Intervalo lematíco
<i>aarónico – abastecer</i>	<i>bazmerreír – bediento,ta</i>	<i>aarónico – abastecer</i>	<i>bazmerreír – bediento,ta</i>
4 de 50	8 de 50	5 de 49	8 de 54
Total = 12 de 100 (12%)		Total = 13 de 104 (12,5%)	

Fonte: autora deste artigo.

Se considerássemos as porcentagens e o número total de verbetes, o DRAE 2001 apresentaria aproximadamente 10,5 mil lemas com problemas na marcação cronológica, e o DRAE 2014 aproximadamente 12 mil. Novamente, as duas edições apresentaram um resultado numérico semelhante no número de verbetes com problemas na marcação cronológica. Esses dados sugerem que a Academia não se valeu de seus recursos informáticos na atualização de seu dicionário, podendo gerar um inchaço macroestrutural desnecessário para a função que o dicionário geral se propõe a cumprir.

Ressaltamos, por fim, que esse tipo de análise quantitativa se baseia num estudo aproximativo, que visa analisar os dados de forma proporcional ao total de entradas no dicionário, considerando que os resultados numéricos obtidos em intervalos lematizados e escolhidos aleatoriamente são muito semelhantes.

b) Análise qualitativa: soluções macroestruturais

Para analisar qualitativamente a macroestrutura do DRAE, julgamos importante observar o tipo de solução proposta pela Academia na disposição dos lemas. Para tanto, seguindo a proposta de Bugueño e Farias (2008), identificamos que a Academia opta por uma solução homonímica de base morfológica. Isto é, registram-se formas homônimas em entradas diferentes por apresentarem morfologias diferentes. Nos intervalos que observamos, identificamos entradas que se referem a um mesmo lexema, mas lematizadas separadamente por sua morfologia, a exemplo de *hebdomada* (substantivo), *hebdomadariamente* (advérbio) e *habdomadario,ria* (adjetivo). Todos possuem o mesmo sentido: são sinônimos de semana (substantivo), semanalmente (advérbio) e semanal (adjetivo).

Por outro lado, o DRAE dispõe seus lemas com um arranjo de entradas de tipo estrutura lisa, isto é, as entradas são organizadas alfabeticamente numa ordenação vertical à esquerda. Embora esse tipo de arranjo facilite a leitura e a identificação dos verbetes por parte do usuário e, quando associada à solução homonímica, propicie uma identificação direta de diferentes formas gramaticais de um mesmo lexema, essa disposição lematizada gera uma alteração relevante no tamanho da macroestrutura do dicionário.

Os verbetes abaixo ilustram um pequeno intervalo do DRAE 2001:

hebillá. (del lat. vulg. **fibella* (...)) f. Pieza de metal o de otra materia, generalmente con uno o vários clavillos articulados en una varilla que la cierra por un lado, los cuales sujetan la correa, cinta, etc. (...)
DRAE (2001)

hebillaje. m. Conjunto de hebillas que entran en un aderezo, vestido o adorno.
DRAE (2001)

hebillar. tr. ant. Poner hebillas o sujetar con hebillas.
DRAE (2001)

hebillero,ra. m. y f. Fabricante o vendedor de hebillas.
DRAE (2001)

hebillleta. f. d. de hebillá.
DRAE (2001)

hebillón. m. aum. de hebillá.
DRAE (2001)

hebilluela. f. d. de hebillá.
DRAE (2001)

Observamos, pois, que esse tipo de arranjo de entradas no DRAE (solução homonímica de base morfológica e estrutura lisa), além de gerar um inchaço na

macroestrutura do dicionário, não aparenta ser mais produtivo que outros arranjos. Portanto, essa disposição lematizada não necessariamente acarreta um ganho para o usuário, uma vez que ao lematizar as variações morfológicas de uma mesma palavra se cria um sistema de remissões, como no caso ilustrado de *bebilla*.

Considerações finais

A partir da análise proposta neste trabalho, pudemos chegar a algumas conclusões quantitativas e qualitativas acerca da macroestrutura do DRAE e da sua atualização.

Em primeiro lugar, no que se refere à análise macroestrutural quantitativa, apesar de a Academia manifestar o embasamento nos registros de seus bancos de dados para a atualização de seu dicionário, os dados sugerem que esse trabalho não se estende a todo o dicionário. Observamos esse fato na forte tendência da Academia em lematizar palavras que não possuem registros (*ghost words*) e reproduzi-las nas edições mais recentes. Tampouco fica claro o critério de inclusão de novos verbetes na última edição, posto que identificamos *ghost words* lematizadas no DRAE 2014 que não estavam incluídas no DRAE 2001.

Com relação à marcação cronológica, também percebemos marcações que não parecem de acordo com os registros dos bancos de dados da própria Academia. Identificamos, também, palavras que deveriam estar marcadas cronologicamente, mas estão lematizadas como palavras da língua atual.

Quanto à análise macroestrutural qualitativa, observamos que a disposição lematizada proposta pela RAE se baseia em uma solução homônima de base morfológica associada a um arranjo de tipo estrutura lisa. Julgamos que esse desenho macroestrutural ocasiona, em grande medida, um inchaço no dicionário, gerando um sistema de remissões que nem sempre acarretam ganho para o usuário.

Por fim, acreditamos que a Academia, enquanto autoridade linguística da comunidade falante de língua espanhola, visa elaborar um dicionário que ofereça uma imagem léxica atualizada da língua espanhola. Contudo, com a análise aqui levada a cabo, julgamos que ainda há muito que avançar no trabalho lexicográfico de atualização do dicionário. Nesse sentido, acreditamos que os recursos informáticos que a Academia elabora e disponibiliza têm muito ainda a contribuir nessa tarefa. A partir do uso dos grandes *corpora* textuais informatizados, é possível proporcionar um dicionário que esteja de acordo com o que um dicionário de língua geral se propõe ser, isto é, cujo objetivo é registrar um momento da língua e que cumpra as necessidades de seus usuários.

Referências

BIDERMAN, M. T. Os dicionários da contemporaneidade: arquitetura, métodos e técnicas. In: OLIVEIRA, A. M. P.; ISQUERDO, A. N. (org.). **As ciências do léxico: lexicologia, lexicografia, terminologia**. Campo Grande: Ed. da UFMS, 1998, p. 129-142.

BIDERMAN, M. T. Análise de dois dicionários do português brasileiro contemporâneo: o Aurélio e o Houaiss. *In: OLIVEIRA, A. M. P.; ISQUERDO, A. N. (org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia.* Campo Grande: Ed. da UFMS, 2004, p. 185-200.

BUGUEÑO, F. V. O que o professor deve saber sobre a nominata do dicionário de língua. **Revista Língua e Literatura**, Frederico Westphalen, v. 6-7, n. 10-11, p. 17-31, 2004/2005.

BUGUEÑO, F. V. Balanço e perspectivas da lexicografia. **Cadernos de Tradução**, Florianópolis, n. 32, p. 15-37, 2013.

BUGUEÑO, F. V. Da classificação de obras lexicográficas e seus problemas: proposta de uma taxonomia. **Alfa**, São Paulo, v. 58, n. 1, p. 215-231, 2014.

BUGUEÑO, F.; FARIAS, V. S. Desenho da macroestrutura de um dicionário escolar de língua portuguesa. *In: XATARA, C.; BEVILACQUA, C. R.; HUMBLÉ, Ph. (org.). **Lexicografia Pedagógica**: pesquisas e perspectivas.* Florianópolis: UFSC/NUT, 2008, p. 129-167.

COSERIU, E. **Lições de linguística geral**. Rio de Janeiro: Ao Livro Técnico, 1980.

DRAE (2001). REAL ACADEMIA ESPAÑOLA. **Diccionario de la Lengua Española**. Madrid: Espasa, 2001.

DRAE (2014). REAL ACADEMIA ESPAÑOLA. **Diccionario de la Lengua Española**. Madrid: Espasa, 2014.

DRAE. REAL ACADEMIA ESPAÑOLA. **Cómo se hace el Diccionario**. Disponível em: <http://www.rae.es/diccionario-de-la-lengua-espanola/como-se-hace-el-diccionario>. Acesso em: 8 abr. 2019.

FARIAS, V. S. **Desenho de um dicionário escolar de língua portuguesa**. 2009. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

FARIAS, V. S. Análise da macro e da microestrutura de dicionários bilíngues português-alemão/alemão-português. **Revista Contingentia**, v. 5, n. 1, p. 76-98, maio 2010.

GRIJELMO, A. **Defensa apasionada del idioma español**. Madrid: Punto de Lectura, 2001.

HAUSMANN, F.; WIEGAND, H. E. Component parts and structures of general monolingual dictionaries: a survey. *In: HARTMANN, R. R. K (org.). **Lexicography**: critical concepts*, v. 3. Londres, Academic Press, 2003.

LARSON, P. Ghost words and new discoveries in the TLIO Old Italian dictionary. **ICHLL International Conference on Historical Lexicography and Lexicology**, 5, Oxford, 16-18 jun. 2010.

MALMKJÆR, K. **The linguistics encyclopedia**. Londres: Routledge, 2004.

SALVADOR, G. **El Diccionario de la Real Academia Española**, 1992. Disponível em: https://cvc.cervantes.es/obref/congresos/sevilla/unidad/ponenc_salvador.htm. Acesso em: 8 abr. 2019.

SVENSÉN, B. Types of dictionaries. *In*: SVENSÉN, B. **A handbook of lexicography: the theory and practice of dictionary-making**. Cambridge: Cambridge University Press, 2009.

WIEGAND, H. E.; BEER, M. Access structure in printed dictionaries. *In*: WIEGAND, H. E. *et al.* (org.). **Dictionaries**. An international encyclopedia of lexicography. Supplementary volume. Berlin: De Gruyter Mouton, 2013.

Anexo – Intervalos lemáticos das letras A e H.

Verbetes	DRAE 2001	DRAE 2014	CORPES XXI	CORDE	Análise
1. Aarónico		✘	0	0	Ghost word
2. Aaronita		✘	0	1 (1967)	Hápax legomena
3. Aba			128	60	✓
4. Abaá	✘		3	0	✓
5. Ababa			2	13	✓
6. Ababillarse			0	0	Ghost word
7. Ababol			2	6	✓
8. Abacá			9	81	✓
9. Abacal			0	0	Ghost word
10. Abacalero,ra			0	0	Ghost word
11. Abacería			19	11	✓
12. Abacero,ra			2 (2001/2004)	1 (1550)	✓
13. Abachar		✘	0	0	Ghost word
14. Abacial			14	41	✓
15. Ábaco			87	60	✓
16. Abacora (ant)			0	0	Ghost word
17. Abacorar			2	1	✓
18. Abad			350	5088	✓
19. Abada			10	5	✓
20. Abadejo	✘		29	91	✓
21. Abadengo,ga	✘		9	315	✓
22. Abadernar	✘		0	0	Ghost word
23. Abadesa	✘		0	909 de 1028 a 1974	✓
24. Abadí	✘		20	0	✓
25. Abadía	✘		581	563	✓
26. Abadiado	✘		0	108	✓
27. Abarquillado,da	✘		11	5	✓

28. Abarquillamiento	x		0	0	Ghost Word
29. Abarquillar			3	0	✓
30. Abarracar			6518	483	✓
31. Abarrado,da			0	2 (1575/1646)	Marca de desusada
32. Abarraganamiento			1	1 (1495)	✓
33. Abarraganarse			2	0	Hápax legomena
34. Abarrajado,da			0	0	Ghost word
35. Abarrajar			0	1 (1535)	Marca de desusada; Hapax legomena
36. Abarramiento		x	0	0	Ghost Word
37. Abarrancadero		x	0	0	Ghost Word
38. Abarrancamiento			0	1 (1957)	Hápax legomena
39. Abarrancar			0	2 (1575)	Marca de desusada
40. Abarrar (desus)			0	3	✓
41. Abarraz			0	2 (1542)	Marca de desusada
42. Abarredera (ant)		x	0	2 (1251)	✓
43. Abarrenar (desus)		x	0	1 (1582)	Hápax legomena
44. Abarrer (desus)		x	0	2 (1490)	✓
45. Abarrisco			0	25	✓
46. Abarrota			1287	4	✓
47. Abarroto			92	2	✓
48. Abarrotería			37	0	✓
49. Abarrotero,ra			18	4	✓
50. Abarse			3	1	✓
51. Abasí			24	1	✓
52. Abastadamente (desus)			0	65	✓
53. Abastamiento (desus)			0	49	✓
54. Abastante (desus)		x	0	33	✓
55. Abastanza (desus)			1	14	Não deveria estar marcada como desusada
56. Abastar			13	128	✓
57. Abastardar		x	0	0	Ghost Word
58. Abastecedor,ra			254	40	✓
59. Abastecer			2673	889	✓
TOTAL LEMAS	50	49			✓

Verbetes	DRAE 2001	DRAE 2014	CORPES XXI	CORDE	Análise
1. Hazmerreír			140	39	✓
2. He ¹⁴			15679	131810	✓
3. He ²			15670	131810	✓
4. Heavy	✘		522	2	✓
5. Heavy metal	✘		270	0	✓
6. Hebdómada			0	13 (1536 – 1623)	Marca de desusada
7. Hebdomadariamente			0	1 (1923)	Hápax Legomena
8. Hebdomadario,ria			7	22	✓
9. Hebén			0	0	Ghost Word
10. Hebetar			0	0	Ghost Word
11. Hebijón			0	0	Ghost Word
12. Hebilla			493	219	✓
13. Hebillaje			0	3 (1878 – 1886)	Marca de desusada
14. Hebillar			10	219	✓
15. Hebillero,ra			1	1	✓
16. Hebilleta			0	12 (1550-1607)	Marca de desusada
17. Hebillón			2	6	✓
18. Hebra			903	564	✓
19. Hebraico,ca			87	318	✓
20. Hebraísmo			4	29	✓
21. Hebraísta			8	6	✓
22. Hebraizante			0	24 (1863-1905)	✓
23. Hebraizar			0	1 (1678)	Marca de desusada
24. Hebreo,a			1526	1629	✓
25. Hebrero ¹			3	1161	✓
26. Hebrero ² (desus) febrero			0	1161 ¹⁵	✓
27. Hebroso,sa			0	1 (1889)	Marca de desusada; Hapax legomena
28. Hebrudo,da			2	0	✓
29. Hecatombe			389	137	✓
30. Hecha			6714	18673	✓
31. Hechiceresco,ca			0	1 (para <i>hechiceresca</i> de 1615)	Hapax Legomena; Marca de

¹⁴ Os *corpora* não permitiram a busca por classe de palavra (uma vez que o lema *he*¹ é um advérbio e *he*² um verbo). Portanto, o índice de frequência se refere às duas formas. Contudo, como o propósito deste trabalho é observar se as palavras apresentam registro nos *corpora*, o alto número de frequência de ambas as formas já justifica sua lematização nos dicionários.

¹⁵ Idem nota de rodapé 13.

					desusada
32. Hechicería			175	269	✓
33. Hechicero,ra			949	1214	✓
34. Hechizante	x		47	0	✓
35. Hechizar			621	144	✓
36. Hechizo,a			1227	945	✓
37. Hecho,cha			96918	138546	✓
38. Hechor, ra			47	84	✓
39. Hechura			416	3562	✓
40. Heciento,ta			0	0	Ghost Word
41. Hectárea			9320	266	✓
42. Héctico,ca			0	5 (1591 – 1606)	Marca de desusada
43. Hectiquez			0	0	Ghost Word
44. Hecto			2	12	✓
45. Hectógrafo			1	1	✓
46. Hectógramo			2	1	✓
47. Hectolítro			62	33	✓
48. Hectómetro			400	11	✓
49. Hectopascal	x		20	0	✓
50. Hectóreo,a			0	7 (1427 – 1618)	Marca de desusada.
51. Hedentino,na			2 (somente forma feminina)	4 (somente forma feminina)	Poderia haver um comentário indicando uso no feminino.
52. Hedentinoso,sa (ant)			0	1 (somente forma feminina)	Hápax Legomena; Poderia haver comentário indicando uso no feminino.
53. Heder			226	138	✓
54. Hediento,ta			2	2 (1494 e 1930)	✓
TOTAL LEMAS	50	54			✓