

O modelo de descrição no dicionário histórico morfológico: o radical de *alto*

ALINA VILLALVA

Universidade de Lisboa

JOÃO PAULO SILVESTRE

Centro de Linguística da Universidade de Lisboa

ANA GUILHERME

Centro de Linguística da Universidade Nova de Lisboa

O conhecimento do léxico do Português ainda tem muitas lacunas, algumas conhecidas, outras insuspeitas. O conhecimento diacrónico, bem como a natureza e relações morfológicas e morfossemânticas das palavras, são domínios em que a lexicografia portuguesa do século XX não conseguiu acompanhar o trabalho desenvolvido para as línguas cognatas, deixando o português em evidente desvantagem na qualidade da descrição, em comparação com outras línguas modernas, como o francês, o castelhano ou o italiano. Apesar de os dicionários geralmente incluírem nos artigos informações sobre a etimologia e a morfologia, estas secções são apresentadas de uma forma descoordenada, que resulta mais da acumulação de notícias recolhidas em dicionários anteriores, do que de uma reflexão sistemática sobre famílias de palavras que se encontram registadas na totalidade da lista de entradas.

O dicionário de radicais que nos propomos desenvolver é um dicionário especializado, que tem como público-alvo os lexicólogos, os lexicógrafos e também os docentes de língua portuguesa. Pretende constituir um modelo de descrição mais circunstanciado, que documente o percurso morfológico e semântico das palavras, mas também avalie o seu uso no português contemporâneo.

A descrição lexicográfica do radical de *alto* será a base para um artigo-modelo com que procuramos explicitar o método de investigação que adotámos.

A constituição de um corpus lexical a partir de dicionários

O foco de análise da atividade lexicográfica é geralmente colocado no número de entradas de um dicionário e na inclusão de mais palavras, por vezes apresentadas como novas. Todavia, esta visão ilude vários factos, que perturbam a descrição do léxico e também o uso dos dicionários.

O número e a qualidade das entradas de um dicionário devem ser definidos em função do uso a que esse dicionário se destina. Tomemos como referência os dicionários de âmbito geral contemporâneo. O que se espera encontrar nestes instrumentos de trabalho é, regra geral, alguma clarificação sobre a ortografia e os sentidos da palavra e, porque só se procuram palavras que já se conhecem, uma confirmação da sua (in)existência. A existência das palavras depende menos da vontade dos dicionaristas e mais do uso dos falantes. A estratégia de incluir todas as palavras que possam em algum momento ter sido dicionarizadas ou atestadas em alguma fonte textual, sem qualquer indicação sobre as características e frequência do uso, tornam os dicionários de âmbito geral em objetos de outra natureza – mais próximos de uma base de dados do que de uma lista de palavras necessárias para uma determinada função. O seu uso como instrumentos de trabalho fica comprometido, por vezes sem que os utilizadores se deem conta de tal facto.

A inovação lexical é geralmente descrita como um movimento de expansão e enriquecimento expressivo das línguas, mas esta expansão pode ser justificada por necessidades de adequação discursiva transitória. É provável que, a longo prazo, novas formações ou empréstimos não sejam incorporados num fundo lexical estável que continue a ser partilhado e reconhecido pela generalidade dos falantes. E se este facto é facilmente observável na língua contemporânea, devemos considerar que também aconteceu em momentos anteriores da história da língua. Os testemunhos históricos de inovação lexical, desde os dicionários aos textos literários, são tradicionalmente considerados como referências *ab quo* para o uso das palavras, quando, na realidade, podem ser monumentos a uma existência efémera, que tem tanto de original quanto de pouco frequente.

Em suma, os dicionários gerais de língua tendem a acumular informação dos dicionários anteriores, sem avaliar o seu uso ou aceitabilidade. Essa acumulação torna-se ainda mais evidente nos dicionários electrónicos, que, num esforço de condensar a informação, não dão elementos sobre restrições de uso¹.

¹ *Dicionário Priberam da Língua Portuguesa* e *Infopedia* são exemplos de dicionários de orientação acumulativa. O *Priberam* é escasso na marcação explícita de registos e restrições ao uso.

Os dicionários são uma fonte privilegiada de informação morfológica e podem constituir uma base documental para a recolha de um corpus de trabalho. Tomemos como exemplo a pesquisa em torno do radical de *alto*. O dicionário de Cândido de Figueiredo, na edição de 1913, regista 340 palavras em que o segmento gráfico *alt* ocorre em posição inicial ou medial². Aceitando as definições propostas pelo lexicógrafo, apenas 97 estão de alguma forma relacionadas com o radical *alt-*. Entre estas, são excluídos os compostos morfológicos, pois apresentam dois ou mais radicais (por exemplo, *planalto*). Assim, do *corpus* abundante de Cândido de Figueiredo restam apenas 41 palavras relacionadas com o radical *alt-*.

A comparação com os dicionários contemporâneos (*Dicionário da Língua Portuguesa Contemporânea/DLPC*, *Infopédia* e *Dicionário Priberam da Língua Portuguesa/Priberam*) traz poucas adições e, sobretudo, revela que poucas palavras têm vindo a ser consideradas fora de uso. Mas o confronto com o *Corpus de Referência do Português Contemporâneo/CRPC* sugere que o estatuto de algumas dessas palavras nos dicionários deve ser objeto de uma marcação específica, pois ou são de ocorrência zero, ou têm uma frequência tão baixa que indiciam a pertença a um registo linguístico delimitável³.

O quadro seguinte assinala as 41 palavras que, a partir da lista de Cândido de Figueiredo, relacionamos com o radical *alt-*, a sua manutenção em dicionários posteriores e os dados de frequência no CRPC.

Figueiredo 1913	DLPC 2001	Infopédia 2014	Priberam 2014	CRPC 2014
alta _{SUBS}	ü	ü	ü	6684
altamente	ü	ü	ü	8211
altanado	û	ü	ü	0
altanar	û	ü	ü	0
altanar-se	û	ü	ü	0
altanaria	ü	ü	ü	14
altaneiramente	û	û	ü	7
altaneiro	ü	ü	ü	43
altania	û	ü	û	0
altar	ü	ü	ü	1417
altareiro	û	û	ü	0
altarista	û	û	ü	0
alteação	û	ü	ü	0

² O dicionário de Cândido de Figueiredo foi digitalizado para servir de base documental ao projeto Dicionário Aberto e está disponível online com ferramentas de pesquisa (<http://www.dicionario-aberto.net>).

³ O *CRPC* tem cerca de 309 milhões de palavras, maioritariamente de textos posteriores a 1970.

Figueiredo 1913	DLPC 2001	Infopédia 2014	Priberam 2014	CRPC 2014
alteado	ü	û	ü	28
alteador	ü	ü	ü	1
alteamento	ü	ü	ü	55
altear	ü	ü	ü	23
alteável	û	ü	û	0
altenado	û	û	û	0
altíssimo	ü	ü	ü	1299
altista	ü	ü	ü	166
altitude	ü	ü	ü	2086
altitudinal	û	ü	û	0
altivamente	ü	û	ü	53
altivar	û	û	û	0
altivez	ü	ü	ü	180
altiveza	û	û	û	0
altividade	û	û	û	0
altivo	ü	ü	ü	174
alto _{ADJ}	ü	ü	ü	50604
altor	û	û	û	0
altosa	û	ü	ü	0
altura	ü	ü	ü	70.768
enaltar	û	û	û	0
enaltecer	ü	ü	ü	675
enaltecimento	ü	ü	ü	50
exaltação	ü	ü	ü	1095
exaltadamente	û	ü	ü	19
exaltado	ü	ü	ü	896
exaltador	û	ü	ü	4
exaltamento	ü	ü	ü	5
exaltar	ü	ü	ü	533
sobreexaltar	ü	ü	ü	0
superexaltado	û	û	û	0

Torna-se importante distinguir as palavras fundamentais e conhecer o seu percurso nos dicionários, no sentido de marcar os casos em que a dicionarização é uma sobrevivência meramente fictícia, resultante mais de opções editoriais do que de estudo lexicológico fundamentado.

Porquê uma organização em radicais

Apesar de o número de palavras de uma língua ser aparentemente elevado, postulamos que a inovação lexical ocorre em torno de um número relativamente restrito de radicais, conjugando processos morfológicos disponíveis com as necessidades semânticas do momento. Se o número de radicais é estável, o mesmo não se pode dizer dos seus significados e do significado que se atribui às diferentes palavras derivadas. Por isso partimos da hipótese de que a base da grande maioria das palavras do português, das mais antigas até às mais recentes, faz uso de um conjunto de radicais que se apresentam sob a forma de palavras simples ou de palavras complexas lexicalizadas.

O nosso objectivo é a identificação desses radicais, e a constituição de uma base de dados que descreva o seu desenvolvimento em português. Não existe uma listagem que reúna simultaneamente informação sobre a evolução morfológica e semântica das palavras portuguesas, nem sequer uma tradição de ordenação do léxico pelas chamadas famílias de palavras⁴.

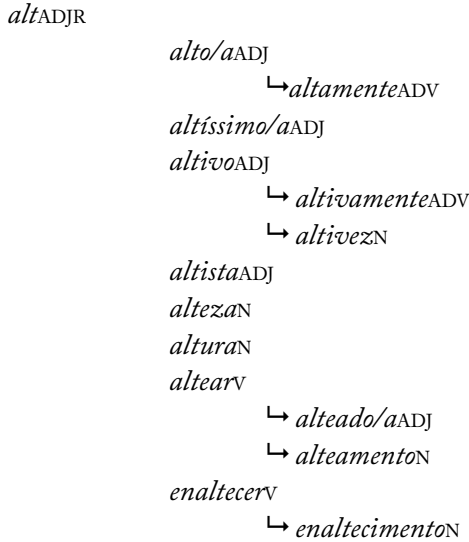
O próprio conceito de família de palavras não é consensual, pelo que os critérios que definem a inclusão de palavras nestes conjuntos poderão diferir⁵. No caso de um trabalho que incide sobre radicais, parece relevante incluir as palavras simples (e.g. *alt-o*), as palavras derivadas (e.g. *alt-ur-a*, *alt-iv-ez*) e as palavras modificadas (e.g. *alt-issim-o*). Em contrapartida, devem excluir-se:

- a. Os empréstimos, pois correspondem a formas lexicalizadas que não podem ser analisadas em português. Por exemplo, *altaneiro* pode ser considerado um empréstimo, pois em português não existe a base **altano*. O mesmo se aplica em palavras decalcadas do latim, que apresentam prefixos e sufixos inexistentes em português, e por isso não analisáveis (e.g. *altitude*, *exaltar*).
- b. Palavras desusadas, que na diacronia foram substituídas por outras formas derivadas (e.g. *altiveza*, desusada; cf. *altivez*).
- c. Palavras cognatas, sem relação morfológica no português (e.g. *alçar vs. altear*, do latim *altiare*);
- d. Palavras que não estão documentadas num corpus textual extenso, apesar estarem dicionarizadas (e.g. *altivar*, *altosa*, *altor*, *altriz*, *alteável*, *alteador*, *alteação*, *enaltar*).

⁴ Há algumas tentativas, em trabalhos que hoje estão desatualizados no que respeita a dados e a opções metodológicas, como o *Dicionário de raízes e cognatos da Língua Portuguesa*, de Carlos Goes (1936) e o *Dicionário morfológico da língua portuguesa* de Evaldo Heckler (1984)

⁵ Ver, a este respeito, Bauer & Nation 1993.

A aplicação dos critérios acima enunciados à descrição morfológica do radical de *alto* pode resumir-se na seguinte estrutura hierárquica:



À primeira vista, esta descrição parece colidir com informação consensualmente aceite e registada em dicionários gerais, que prometem aos seus leitores informação etimológica e derivacional.

Nestes dicionários confundem-se conceitos como radical (que refere a parte de uma palavra que é partilhada pelos elementos pertencentes a uma determinada família de palavras), étimo (que é a forma a partir da qual uma outra palavra provém) e empréstimo (que é uma inovação lexical a partir do contacto com outra língua). Por serem apresentados com notória disparidade de critérios, nos campos da “origem” encontram-se informações aparentemente contraditórias. Sobre *enaltecer*, por exemplo, três dicionários propõem diferentes análises:

Espanhol *enaltecer* (1607) ‘id.’, de alto; ver alt- (*Dicionário Houaiss* 2009)

Espanhol *enaltecer* (*Príberam*)

De en+alto+ecer (*Infopédia*)

Os dois primeiros indicam que a palavra resulta de um empréstimo (com menção da data da primeira atestação, no caso do dicionário Houaiss 2009, mas a *Infopédia* apresenta uma análise morfológica que atribui à palavra uma estrutura parassintética. Pensando em palavras como *enriquecer*, *enfraquecer*, *envelhecer*, a explicação poderia ser congruente. Todavia, *enaltecer* apenas está documentado em *corpora* de português a partir do século XIX, enquanto as outras palavras

ocorrem desde o século XIII; em castelhano, a palavra equivalente está registada com frequência desde o século XV⁶. Ou seja, a diacronia confirma que se trata uma palavra recente, muito provavelmente resultante de um empréstimo, que se adapta ao português apoiada na analogia com formas pré-existentes.

A informação diacrónica permite ainda identificar as variantes morfológicas, que tiveram um uso episódico na história da língua ou vieram substituir por completo uma forma anterior. Na família de *alto*, é de referir o par *altivez* e *altiveza*. *Altivez* é a forma mais antiga e é frequente até ao português contemporâneo, ao passo que *altiveza* apenas está documentada, e com baixa frequência, entre os finais do século XVI e o início do século XVIII. Em castelhano a forma *-eza* está documentada desde o início do século XV, pelo que é admissível alguma interferência. De qualquer modo, o período em que foram possíveis coocorrentes e equivalentes terminou há quase três séculos, com vantagem para a forma anterior, que é a única desde o português moderno. O facto de ainda estar registada no *Vocabulário Ortográfico do Português* pode indiciar uma contemporaneidade que é perfeitamente equívoca.

A organização de um dicionário histórico morfológico

A construção de um dicionário histórico morfológico, organizado a partir dos radicais, relaciona-se com a constituição de um vocabulário fundamental do português contemporâneo, em que as fronteiras entre a língua em uso e a memória da língua se tornam mais claras.

Este trabalho de síntese depende de esforços a montante, nomeadamente a identificação das primeiras atestações lexicográficas e o roteiro diacrónico da dicionarização. Na posse desses dados, os *corpora* de textos literários e de textos metalinguísticos permitem confirmar as datações lexicográficas e localizar testemunhos do uso efetivo das palavras. Estas informações são essenciais para esclarecer a formação de palavras em português numa perspectiva diacrónica, pois permitem identificar a acumulação acrítica de palavras com significados que se sobrepõem.

Neste momento, trabalhamos na definição da lista de radicais e na construção de um modelo de descrição lexical. No que respeita aos radicais, começámos por ensaiar um modelo de descrição de adjetivos (*bravo, esquisito, longo, comprido*) (Villalva & Silvestre (2011), Silvestre & Villalva (2014), Villalva & Silvestre (no prelo). Seleccionámos palavras com um longo percurso diacrónico e com

⁶ De acordo com as bases de dados *Corpus Lexicográfico do Português*, *Corpus do Português* e *Corpus del Nuevo Diccionario Histórico del Español*.

origem interlinguística. Estes casos permitem análises contrastivas e geralmente revelam processos de mutações semânticas e percursos derivacionais distintos.

Do estudo do radical de *alto* resultou um modelo de descrição que contempla a informação etimológica, a amplitude semântica, as relações de derivação e a datação conhecida das formas.

A etimologia procura recuar na história do radical somente para explicar a forma existente na língua portuguesa, com o significado relevante para a compreensão da semântica desse radical em português. A amplitude semântica pode incluir as acepções que se desenvolvem na diacronia. A sua ordenação subordina-se à data das formas mais antigas e não à frequência ou a uma avaliação do que pode ser percebido como um sentido mais geral.

L. *altu*-ADJ

(de *altu*- ‘alimentado, crescido’, participio de *alo*, -ere ‘alimentar’)

1. profundo	13 – 18
2. alto	13 –
3. ilustre	14 –

1. profundo

<i>alto</i> /aADJ	profundo: 13 - 18
<i>altíssimo</i> /aADJ	muito profundo: 17
<i>altamente</i> ADV _{loc}	profundamente: 15 - 17
<i>alturas</i> SUBS	profundidade: 13 - 19
<i>alto</i> SUBS	mar profundo: 14 -

2. alto

<i>alto</i> /aADJ	grande: 13 –
↳ <i>alto</i> ADV _{modo}	elevado (som): 16 –
<i>altíssimo</i> /aADJ	muito grande: 19 –
<i>alto</i> SUBS	cimo: 15 –
<i>alta</i> SUBS	cimo: 13 –
	subida: 19 –
↳ <i>artista</i> ADJ	em subida: 20 –
<i>alta</i> SUBS	permissão: 18 –
<i>alteza</i> SUBS	qualidade do que é alto: 14,18
<i>alturas</i> SUBS	qualidade do que é alto: 15 –
<i>altear</i> V	tornar mais alto 19 –

3. ilustre

<i>altíssimo</i> /aADJ	muito ilustre: 17 –
------------------------	---------------------

↳ <i>altíssimo</i> _{SUBS}	deus: 15 –
<i>altamente</i> _{ADV} _{modo}	de um modo ilustre: 14 –
<i>altivo</i> _{ADJ}	superior (positivo): 16 –
↳ <i>altiveza</i> _{SUBS}	qualidade do que é superior (positivo): 17
↳ <i>altivez</i> _{SUBS}	qualidade do que é superior (positivo): 17 –
<i>alteza</i> _{SUBS}	qualidade do que é ilustre: 14 – 18
↳ <i>alteza</i> _{SUBS}	pessoa ilustre: 15 –

Esta ordenação diacrónica dos factos morfológicos e semânticos oferece aos lexicógrafos uma fonte de informação que presentemente não existe sistematizada. Apesar de trabalharmos a partir de *corpora* em alargamento, e por isso sujeitos a retrodatações à medida que novas fontes são adicionadas, é possível desde já estabelecer um quadro coerente de descrição no que respeita ao aparecimento de formas complexas.

A identificação do nexos temporal é dirimente para comprovar o desuso ou a substituição de palavras. Pode fornecer dados fiáveis para uma revisão da informação sobre morfologia dos dicionários do português contemporâneo, bem como para tornar os dicionários bilingues mais adequados a tradutores e estudantes, eliminando palavras fora de uso que continuam a exigir equivalentes.

Referências bibliográficas

- Bauer, L. & Nation, P. (1993). Word Families. *International Journal of Lexicography*, 6 (4) 253-279.
- Corpus de Referência do Português Contemporâneo*. Centro de Linguística da Universidade de Lisboa. Online: <http://www.clul.ul.pt/pt/recursos/183-crpc#cqp> [10/09/2014].
- Corpus do português: 45 million words, 1300s-1900s*. Online: <http://www.corpusdoportugues.org> [10/09/2014].
- Corpus Lexicográfico do Português*. Centro de Linguística da Universidade de Lisboa – Universidade de Aveiro. Online: <http://clp.dlc.ua.pt> [10/09/2014].
- Figueiredo, C. de (1913). *Novo dicionário da língua portuguesa*. Porto: Typ. da Empr. Litter. e Typographyca
- Goes, C. (1921). *Dicionário de raízes e cognatos da língua portuguesa*. Belo Horizonte: Paulo Azevedo & Cia.
- Heckler, E., Back, S., Massing, E. R. (1984-1988). *Dicionário morfológico da língua portuguesa*. São Leopoldo: Unisinos.
- Houaiss, A. & Villar, M. (2009). *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Objetiva.

- Infopédia. Dicionário da Língua Portuguesa da Porto Editora.* Online: <http://www.infopedia.pt/lingua-portuguesa/> [10/09/2014].
- Le Trésor de la langue française informatisé.* Online: <http://atilf.atilf.fr/tlf.htm> [10/09/2014].
- Nuevo tesoro lexicográfico de la lengua española.* Online: <http://ntlle.rae.es/ntlle/SrvltGUILoginNtlle> [10/03/2014].
- Silvestre, J. P., Villalva, A. (no prelo). Mutations lexicales romanes: esquisito, bizarro et comprido. *InnTrans: Innsbrucker Beiträge zu Sprache, Kultur und Translation.* Frankfurt am Main: Peter Lang Verlag.
- Silvestre, J.P. & Villalva, A. (2014). Morphological Historical Root Dictionary for Portuguese. In Abel, A., Vettori, C. & Ralli, N. (eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* Bolzano/Bozen: EURAC research, 967-978.
- Villalva, A., Silvestre, J. P. (2011). De bravo a brabo e de volta a bravo: evolução semântica, análise morfológica e tratamento lexicográfico de uma família de palavras. *ReVEL* 9, 17. Online: http://www.revel.inf.br/files/artigos/revel_17_de_bravo_a_brabo.pdf [10/09/2014].
- Vocabulário Ortográfico do Português.* Online: <http://www.portaldalinguaportuguesa.org/vop.html> [10/03/2014].