

The Lexicography of Portuguese

Ana Frankenberg-Garcia

Abstract

This chapter deals with the lexicography of Portuguese. Despite the geographical spread of Portuguese, only Brazil and Portugal have a lexicographic tradition. Both these traditions are presented, and briefly contrasted, with a focus on the lexical differences, the role of the language academies, the latest spelling reform, and an overview of recent paper and digital dictionaries.

Introduction

The Portuguese language has its origins in Galicia and the North of Portugal. Like other Romance languages, it is an inflectional language and the majority of its words are derived from Latin. However, a considerable number of Arabic words were incorporated into Portuguese with the Moorish occupation of the Iberian Peninsula in the Middle Ages, and the Portuguese seafaring tradition in the Age of Discoveries brought in loans from all over the world. More recently, new technologies and globalization have opened the doors to the incorporation of many more loans, particularly from English.

Today, with around 260 million native speakers, Portuguese is the official language of Angola, Brazil, Cape Verde, East Timor, Guinea-Bissau, Mozambique, Portugal and São Tomé and Príncipe. It is also one of the official languages (together with Chinese) of Macao, and is spoken in several other countries where there are significant Portuguese-speaking immigrant communities. Despite this geographical diversity, only Brazil and Portugal have a tradition of lexicography.

Description

First Portuguese dictionaries and the emergence of Brazilian dictionaries

The first proper dictionary for Portuguese, the [*Vocabulário Portuguez e Latino, Áulico, Anatômico, Architectonico*](#), was a bilingual Portuguese-Latin dictionary compiled by the Catholic priest Rafael Bluteau between 1712-1728 (Verdelho 2002, Biderman 2003). General monolingual dictionaries would only emerge in the late eighteenth century, with the publication of a number of works, the most influential of which was António de Morais Silva's [*Diccionario da Lingua Portuguesa*](#) in 1789. Both these dictionaries have been digitized and aligned with their printed images by the Brasileira USP digital library at the University of São Paulo.

The *Morais* dictionary was initially based on the Portuguese section of Bluteau's bilingual Portuguese-Latin dictionary, which was then revised and updated in 1813 and 1823, and continued to be revised and updated after *Morais's* death until its 10th and last edition edited by José Pedro Machado and renamed the *Grande Dicionário da Língua Portuguesa* came out in twelve volumes published between 1949 and 1959. Although *Morais* was Brazilian born, the first edition of his dictionary was published in Portugal at a time Brazil was still a Portuguese colony. Throughout its lifespan, the *Morais* continued to be an essentially European Portuguese dictionary.

The lexicography of the two countries began to diverge over one hundred years after Brazil's independence from Portugal, with the publication of the *Pequeno Dicionário Brasileiro da Língua Portuguesa* (PDBLP) in 1938. According to Biderman (2003:58), this small but hugely successful volume was the first that began to record Brazilian words and phrases that are not commonly in use in Portugal. However, after its eleventh edition in 1967, the publishing house that had launched the PDBLP was closed down by the military, at a time Brazil was under the rule of a military dictatorship. Then in 1975, the former chief editor of the PDBLP, Aurélio Buarque de Hollanda Ferreira, published the first edition of the *Novo Dicionário da Língua Portuguesa*, which came to be known as the *Aurélio*. For a long time, this was the only truly Brazilian Portuguese language dictionary available.

Lexical differences between European and Brazilian Portuguese

The lexical differences between Brazilian and European Portuguese date back to the times when Brazil was still a colony. Many new words were incorporated into Brazilian Portuguese from Kikongo, Kimbundu and Umbundu when slaves speaking Bantu languages were brought over from Africa to work in the Brazilian plantations. *Bagunça* 'confusion', *moleque* 'boy' and *quitanda* 'grocer's' are just a few of the hundreds of words that remain even today part of everyday Brazilian Portuguese. In addition to words coming from Africa, there are words in Brazilian Portuguese which were already there when the Portuguese explorers arrived, since they derive from native indigenous languages like Tupi. *Cupim* 'termite', *catapora* 'chicken pox' and *siri* 'crab' are examples of words that come from Tupi and are used in Brazil but not in Portugal.

Further lexical differences emerged later, as Portugal seemed more open to loans from French like *rés do chão* from *rez-de-chaussée* 'ground floor' and *betão* from *beton* 'concrete', while Brazil tended to prefer the respective English loans *térreo* and *concreto*. Other differences arose as new words were simply coined differently on either side of the Atlantic. For example, *goal* became *gol* in Brazilian Portuguese (BrP) and *golo* in European Portuguese (EuP), while *file* became *arquivo* in BrP and *ficheiro* in EuP, and *mobile phone* became *celular* in BrP and *telemóvel* in EuP. A fascinating account of how Brazilian Portuguese evolved is available at the [Museu da Língua Portuguesa](#) 'Portuguese Language Museum' in São Paulo, which hosts a wealth of information about the language in engaging hands-on interactive displays.

Although this is not the place for an extensive analysis of the differences between Brazilian and European Portuguese, from a lexicographic point of view it is important to note that apart from lexis there are also spelling variations affecting the two varieties of Portuguese, such as words like *facto* (EuP) and *fato* (BrP) 'fact', *recepção* (EuP) and *recepção* (BrP) 'reception' and *bilingue* (EuP) and *bilíngue* (BrP).

European and Brazilian Portuguese also differ with regard to pronunciation, some aspects of grammar and a few collocations, but as shall be seen, these differences have not had as much impact on Portuguese language dictionaries simply because traditionally they tend to focus on meaning alone.

Language academies and vocabulary lists

Portugal and Brazil have two separate language academies. Portugal's [Academia de Ciências de Lisboa](#) was founded in the late eighteenth century at a time when other academies were emerging in Europe. Its mission is to act as an advisory body for the Portuguese government in language matters, but its influence has not been very significant. Despite its over 200 years of existence, the Academia de Ciências de Lisboa only managed to publish a complete dictionary in 2001, *The Dicionário da Língua Portuguesa da Academia de Ciências de Lisboa*.

Over one hundred years after the foundation of the Academia de Ciências de Lisboa, the [Academia Brasileira de Letras](#) was established in Rio de Janeiro in 1897 by a group of prominent Brazilian literary figures under the leadership of Machado de Assis, often regarded Brazil's greatest writer. Like its Portuguese counterpart, it is a consultative body only. The only dictionary published by the Academia Brasileira de Letras to date is the *Dicionário Escolar de Língua Portuguesa*, a small dictionary for students published in 2009, which is not as nearly as influential as Brazil's commercial *Houaiss* and *Aurélio* dictionaries. Note, however, that the two latter were authored by two prominent members of the Brazilian Academy (Antônio Houaiss and Aurélio Buarque de Hollanda Ferreira, respectively).

The most significant publication of the Academia Brasileira de Letras is probably the [Vocabulário Ortográfico da Língua Portuguesa](#), a list of 381 thousand lemmas recognized to be part of Brazilian Portuguese. The fifth edition of the Vocabulário Ortográfico published in 2009 has become especially important in Brazil because of the orthographic changes that have taken place with the recent implementation of the 1990 Spelling Reform. Although the Vocabulário Ortográfico is not a proper dictionary in that it is simply a list of words along with their grammatical categories (it does not contain any information on meaning or use), it has come to be adopted by lexicographers and the public in general as a standard reference regarding how words should be spelled in Brazilian Portuguese.

For European Portuguese, an analogous list of 211 thousand lemmas with their respective part-of-speech categories and inflections — the [Vocabulário Ortográfico Português](#) — has been put together by Portugal's [Instituto de Linguística Teórica e Computacional](#) (ILTEC). In its Portal da Língua Portuguesa 'Portuguese Language Portal', ILTEC also lists separately the [words that have changed with the 1990 Spelling Reform](#) and enables users to download a free tool for converting old spelling into new spelling, the [Lince](#) converter. The Vocabulário Ortográfico has become an essential reference and is nowadays far more influential in Portugal than the dictionary published by the Academia de Ciências de Lisboa in 2001.

The 1990 spelling reform

The [1990 spelling reform](#) of the Portuguese language, which has only recently come into effect in Portugal and Brazil and is still in a transitional implementation phase in both countries, has eliminated many but not all the differences in spelling between the two standard varieties of the language. While the words exemplified in the section on lexical differences above remain distinct for phonetic reasons, many words that are pronounced in the same way but were spelt

differently before the reform have now been unified. Thus words like *lingüística* (BrP)/*linguística* (EuP) 'linguistics' and *idéia* (BrP)/*ideia* (EuP) are now respectively spelt without the umlaut and the acute accent in Brazil, and words like *acto* (EuP)/*ato* (BrP) 'act' and *ótimo* (EuP)/*ótimo* (BrP) 'fine' are now respectively spelt without the silent C and the silent P in Portugal.

The spelling reform has led Portuguese dictionaries published after 2009 on both sides of the Atlantic to invest in new editions with the updated orthography. However, the new spelling has not been straightforward to implement in dictionaries. To begin with, the reform affects the alphabetical order of certain headwords. Although the mere removal of an accent, like the change from *idéia* to *ideia*, does not affect the A-Z order, other changes such as the removal of a silent consonant, like the change from *ótimo* to *ótimo*, or the removal of a hyphen and duplication of the linking consonant, like the change from *contra-revolução* to *contrarrevolução* 'counter-revolution', have resulted in changes in the alphabetical listing of headwords. In the present transitional phase, users still familiar with the old spellings may find it difficult to retrieve words that now appear in a different place in their dictionaries if this is not signaled to them.

The current situation is that some dictionaries list old spellings as headwords with cross-references to the corresponding new-spelling headwords, some have simply opted to remove old spellings altogether, and some have included a note referring back to the old spelling under the new spelling entry. While the latter does not appear make a lot of sense (why would a user looking up a word under the new spelling want to know about the old spelling?), the decision to show old spellings nested under new spellings may have to do with the fact that there are a number of people, particularly in Portugal, who are against the spelling reform and make a point of continuing to use old spellings.

It is of course much easier to deal with old-to-new-spelling cross-references in electronic than in print editions, where the alphabetical order is not an issue and users looking up old spellings can be automatically redirected to the new spelling. However, of the six electronic dictionaries examined in 6.2, only three actually redirect users looking up old spellings to new spellings (the Priberam, the Porto Editora and the Aurélio). When looking up an old spelling in the other three dictionaries (the Michaelis, the Aulete and the Houaiss), the user will not be able to retrieve any results.

The spelling reform has also affected the treatment of compounds in Portuguese dictionaries. Traditionally, apart from single words, only hyphenated compounds like *guarda-chuva* 'umbrella' used to figure as separate headwords. However, with the spelling reform, hyphenated compounds linked by prepositions like *maçã-de-adão* 'adam's apple' are now spelt without the hyphen (*maçã de adão*). This means in some dictionaries the compound has disappeared from the headword list because it is no longer hyphenated, while in others this has created a problem of inconsistency, for compounds that used to be hyphenated continue to be listed as main headwords, whereas similar compounds like *maçã do rosto* 'cheekbone', which were not hyphenated before the spelling reform, continue to appear nested under the first lexical word.

In Portugal, where the spelling reform has encountered more resistance by certain sectors of the population, a flexible approach to the new vs old spelling conundrum has been adopted by the Priberam and the Porto Editora dictionaries, where first-time users are prompted in a pop-up window to choose whether they prefer the old spellings or the new ones, and these settings are then saved in cookies for future look-ups.

A snapshot of monolingual Portuguese dictionaries published after 2000

In this section I shall give a brief description of eight major Portuguese monolingual dictionaries published after 2000, five of which were published in Brazil and three in Portugal. Although there are other Portuguese monolinguals available, the eight dictionaries listed in Table 1 are probably the most influential ones in terms of monolingual Portuguese lexicography in the twenty-first century.

Table 1. Major monolingual Portuguese dictionaries published after 2000 in Brazil and Portugal

Country of publication	Name (short name in bold)	Current edition/year	Formats
Brazil	<i>Novo Dicionário Aurélio da Língua Portuguesa</i>	4th/2009	<ul style="list-style-type: none"> • Print • CD-ROM • Simplified free online interface at http://www.dicionariodoaurelio.com/
	<i>Dicionário Houaiss da Língua Portuguesa</i>	1st/2009	<ul style="list-style-type: none"> • Print • CD-ROM
	<i>Michaelis Moderno Dicionário da Língua Portuguesa</i>	1st/2004	<ul style="list-style-type: none"> • Print • CD-ROM • Free online interface at http://michaelis.uol.com.br/moderno/portugues/index.php
	<i>Dicionário UNESP do Português Contemporâneo</i>	1st/2004	<ul style="list-style-type: none"> • Print
	<i>Novíssimo Aulete – Dicionário Contemporâneo da Língua Portuguesa</i>	1st/2012	<ul style="list-style-type: none"> • Print • Free online interface at http://aulete.uol.com.br/
Portugal	Dicionário da Língua Portuguesa Contemporânea da Academia de Ciências de Lisboa	1st/2001	<ul style="list-style-type: none"> • Print
	Dicionário Editora da Língua Portuguesa (Porto Editora)	not specified /2013	<ul style="list-style-type: none"> • Print • Free online interface at http://www.infopedia.pt/lingua-portuguesa/
	Dicionário Priberam da Língua Portuguesa	1st/2008 -2013	<ul style="list-style-type: none"> • Free online interface at http://www.priberam.pt/dlpo/ • Available as MS Word supplement • Applications for Android, Windows Phone and iOS • Integration with Kindle and Kobo

Academic print dictionaries

Of the dictionaries listed in Table 1, the Academia and the UNESP ones are very different from the rest. The two are the product of academic rather than commercial initiatives and they are the only ones to have been originally developed with recourse to Portuguese language corpora.

The Academia dictionary was funded by the Portuguese Government and is the only dictionary in Table 1 produced under the auspices of an officially recognized language planning institution – the Academia de Ciências de Lisboa, whose previous attempts to publish official dictionaries for the Portuguese language in 1793 and in 1976 never went beyond the letter A (Verdelho 2002). Unlike any previous Portuguese language dictionary, some of the information in the Academia dictionary is corpus-based. The corpus used was the [Corpus de Referência do Português Contemporâneo](#) (CRPC), developed at the University of Lisbon and available online at the link above. The CRPC contains around 310 million written words and 1.6 million spoken ones from a variety of texts dating from the nineteenth century to 2006. Most of its texts are from Portugal, but there are also texts from Brazil, and other Portuguese-speaking regions in the world. For Biderman (2003:64), the Academia dictionary cannot be considered a dictionary of contemporary Portuguese as stated in its title, for the reference corpus used was diachronic. From the information available, it is not clear what the composition of the corpus was back in the nineties when the Academia dictionary was written, and neither it is clear how exactly the corpus was used in the dictionary compilation process. There is no indication that any frequency data was used, and in the preface to the dictionary (p. xiv) the authors acknowledge that in addition to the corpus they also used information from other previously existing dictionaries. The examples given in the Academia dictionary come from citations by renowned authors, which is classical in Portuguese lexicography, and also from a mix of corpus and invented examples illustrating usage. The main differences between the Academia dictionary and traditional Portuguese monolinguals are that (a) it contains more examples than usual; (b) some information on collocation and word grammar (helping language production and not just reception) is given, although this information is embedded in examples rather than explicitly highlighted; and (c) it provides phonetic transcriptions.

The UNESP dictionary – developed by academics at Brazil’s Universidade Estadual Paulista - is the only truly modern, corpus-based Portuguese language dictionary in table 1. A 90-million-word corpus compiled at the university’s Laboratório de Lexicografia ‘Lexicography Laboratory’ and containing a variety of texts published in Brazil from the 1950s onwards was initially used to determine the headwords of the dictionary, which were selected on the basis of frequency data. By the time the dictionary was published in 2004, the corpus used had doubled in size, although it is still very small for today’s standards. Unlike traditional dictionaries, the UNESP dictionary is the only Portuguese dictionary to treat multiword units (including unhyphenated compounds like *mão única* ‘one-way (of a road)’ and fixed expressions like *mão na roda* ‘help’) as separate headwords. Another feature that distinguishes the UNESP dictionary is that, like the Academia dictionary in Portugal, it aims to help users not just with language reception but also with language production. In this respect, there is some information on word frequency and there is explicit information on the prepositions that need to be used after verbs followed by indirect objects, such as *basear (+em)* ‘to base (+on)’.

Notwithstanding the importance the Academia and the UNESP dictionaries have had in terms of recent lexicographic advances in Portugal and Brazil, unfortunately both these dictionaries exist only in print are not widely used in the Portuguese-speaking world. The Academia dictionary never went past its first edition, which has been out of print for many years, and the UNESP dictionary is still in its first edition despite the fact that it was published ten years ago.

Commercial e-dictionaries

The other six dictionaries in table 1 are commercial dictionaries that are widely used in Brazil and Portugal today. They are all available in both print and electronic form, except the Priberam dictionary, which does not have a paper edition, but started off as the online interface to the printed dictionary *Novo Dicionário Lello da Língua Portuguesa*. In the brief outline that follows, I shall be referring specifically to the CD-ROM editions of the Houaiss and the Aurélio, and to the online interfaces to the Michaelis, the Aulete, the Porto Editora and the Priberam dictionaries. (Note that while the CD-ROM editions are fixed, the February 2014 online versions described here may have been subjected to changes after the publication of the present chapter.)

Although these commercial e-dictionaries are all very different from one another, what they have in common is that their actual content is still very much influenced by traditional Portuguese lexicography and the conventions of print editions, focusing essentially on meaning.

All dictionaries have made efforts to update their headword list with words that are relatively new in Portuguese and to add new senses to already existing words, but corpus analyses suggest that there are still many gaps to be filled in. Established new words like *roteador/router* 'router' do not appear in all dictionaries. Only the Porto Editora dictionary lists *tuitar* 'tweet' in the sense that it is used today, while the Michaelis and the Aulete point to a meaning that is no longer in use 'to protect' and the other three dictionaries do not list the word at all. Moreover, in all six dictionaries there are many words that are hardly ever used in the language anymore - such as *fradice* 'relative to friar' - and which have neither been removed from the headword list nor marked as dated.

From a technological viewpoint, some electronic editions are more advanced than others. As in any e-dictionary, the headwords can be searched by typing them into a search box. In the Aurélio, the Houaiss the Priberam and the Porto Editora, a moveable alphabetical list is shown as each letter of a word is typed in. In the Michaelis and the Aulete, words written in the search box can only be retrieved after typing them out in full and clicking Enter. If a word is misspelled, the Priberam and the Porto Editora automatically provide similar spelling alternatives to help the user. The latter also opens up a dialogue box where the user can suggest the addition of a new entry to the dictionary. The Houaiss prompts the user to check whether the Caps Lock key is active (it does not handle upper case) or to try an advanced search for similarly spelled words. The Aurélio prompts the user to recheck his/her spelling or to try another type of search. The Aulete and the Michaelis simply return a message stating the word has not been found. As mentioned in the introduction, Portuguese is an inflectional language, but if the user types in an inflected form, only the Priberam and the Porto Editora are capable of retrieving the corresponding lemma.

If the user enters a non-hyphenated compound, the Priberam is able to retrieve it even when it is nested in an entry. The Aurélio retrieves automatically those which figure as separate headwords, but to retrieve nested compounds the user needs to tick the option for searching within the full text of the dictionary. The Porto Editora and the Aulete are able to retrieve compounds that are listed as headwords, but not those which are embedded in an entry. Users cannot search for non-hyphenated compounds in the Michaelis, and the Houaiss does not contain any non-hyphenated compounds at all.

The spelling conventions of Portuguese make the pronunciation of words fairly predictable, and phonetic transcriptions are not usually provided in Portuguese language dictionaries (the Academia dictionary in 6.1 being a notable exception). Not surprisingly, none of the six e-dictionaries examined here include phonetic transcriptions. There are also no links to sound files where users can listen to the pronunciation of headwords. No information on syllabication is given either, although the widespread use of word processors, whereby words that do not fit on a line are automatically shifted to the next line, has made syllabication irrelevant to most people today.

Apart from information on part of speech, little other grammatical information is given except occasional notes on capitalization (in Houaiss, Aurélio and Porto Editora) and irregular plural and feminine inflections (all except Porto Editora). However, when the headword is a verb, the Aurélio, the Porto Editora and the Priberam dictionaries also contain links to verb conjugation tables.

In Portuguese, there are many lemma forms like *azul* 'blue' that can be classified either as adjectives or nouns, depending on the context of use. In all six dictionaries such words belong to the same entry. Sometimes the different part-of-speech categories are listed separately (i.e., first noun then adjective or vice-versa), sometimes they are merged (i.e., noun and adjective).

With regard to the ordering of senses, irrespective of frequency, in all six dictionaries concrete senses tend to be presented first, followed by abstract senses, then figurative senses, then senses used in specialized domains. The Porto Editora, the Michaelis and the Aurélio then list compounds followed by idioms at the end of the entry, whereas the Aulete and the Priberam list compounds and idioms together but sorted in alphabetical order. The Houaiss does not list any compounds or idioms at all.

With regard to context of use, the Aulete, the Aurélio and the Houaiss attempt to provide supporting examples that indirectly show which prepositions must follow certain words. The Aulete is the one that best shows this information by underlining in the examples given both the target word in question and its associated preposition. The Priberam contains a few examples too, but they are mostly examples illustrative of meaning rather than use. The number of examples given in those dictionaries is very limited considering e-dictionaries do not have the space restrictions of print editions, and the examples themselves seem to be a mix of invented examples and citations from famous authors rather than corpus-based. The Michaelis provides haphazard and merely illustrative examples in some entries, and no examples at all are given in the Porto Editora. Interestingly, despite not normally providing examples within the entry itself, the Priberam dictionary lists beneath each entry automatically generated snippets of the word being used in online news, blogs and tweets of the day (note, however, that for obvious reasons this functionality does not work for rarer words that have not been in use on the Web).

From a visual perspective, dictionary senses and definitions are presented very compactly all on the same line and in black only in the Michaelis, as if it were a print edition with severe limitations of space and wishing to save money on color. The five other dictionaries have opted for a one-sense-per-line presentation and color coding different types of information such as examples, idioms and metalinguistic data. Additionally, in the Priberam, links to images are provided for some headwords and senses.

The future

Apart from the Michaelis, which appears to be no more than a static electronic replica of its print edition, it seems that the five other commercial e-dictionaries examined in 6.2 are making valiant efforts to make the most of the new medium. The Priberam seems to be particularly advanced in this respect, and apart from the sophisticated interface to the online dictionary itself, it also offers applications for Android, Windows Phone and iOS and integration with Kindle and Kobo. There is still however plenty of room for future development in Portuguese lexicography.

In addition to addressing the technological and content shortcomings pointed out in the previous sections, Portuguese dictionaries of today would benefit enormously from using corpus data in order to incorporate improvements such as giving users more information about frequency, usage and collocation, determining which new headwords to include and which disused headwords to leave out or mark as dated, updating the definitions of words that have acquired new meanings, making informed decisions on the ordering of different part-of-speech categories and senses within headwords, providing good dictionary examples that would help users with language comprehension and production, and so on.

As referred to in 6.1, the Academia dictionary published in 2001 seems not to have taken full advantage of the Corpus de Referência do Português Contemporâneo which was available to its team of lexicographers at the time it was compiled. The UNESP dictionary which then came out in 2004 was based on a relatively small (for lexicographic purposes) 90-million-word corpus of Brazilian Portuguese. There are now much larger corpora of contemporary Portuguese language corpora available. The two biggest ones are the Corpus Brasileiro and the ptTenTen corpus.

The Corpus Brasileiro, developed by a group of researchers led by Tony Sardinha at the Catholic University of São Paulo contains over one billion tokens from a variety of genres in Brazilian Portuguese (see <http://www.sketchengine.co.uk/documentation/wiki/Corpora/Brasileiro> for details). It can be searched via the AC/DC interface (Santos and Bick 2000) at the [Linguateca Portal](#) and via the [SketchEngine](#) (Kilgariff et al 2004).

The ptTenTen corpus (Kilgariff et al 2012), also available from the [SketchEngine](#), contains over 3 billion tokens of European and Brazilian Portuguese from the web (around 1.5 billion each), with unwanted material and duplicates removed, then processed with the PALAVRAS parser (Bick 2000) and post-processed by Pete Whitelock to optimize Word Sketch output. The fact that the corpus outputs Word Sketches, i.e., automatic corpus-based accounts of a word's grammatical and collocational behavior (Kilgariff et al 2004), in both Brazilian and European Portuguese, make it a very powerful tool for the future development of Portuguese lexicography.

Other welcome improvements include links to sound files showing how words are pronounced in Brazilian and European Portuguese, which can differ quite substantially, more links to images (only the Priberam contains a few images so far) and integrated thesauri.

References

- Acordo Ortográfico da Língua Portuguesa — 1990* 'Portuguese Spelling Reform of 1990'. Available from <http://portal.mec.gov.br/arquivos/pdf/acordoortografico.pdf>
- Bick, E., 2000, *The Parsing System Palavras, Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*, Aarhus University Press.
- Biderman, M.T., 2003, 'Dicionários do português: da tradição à contemporaneidade', *Alfa, Revista de Linguística*, vol. 47 no.1, Available from <http://seer.fclar.unesp.br/alfa/article/view/4232/3827>
- Kilgarriff, A., P. Rychlý, P. Smrz and D. Tugwell, 2004, 'The Sketch Engine', *Proceedings of EURALEX 2004*. Lorient, France.
- Kilgarriff, A., Pomikalek, J., Jakubíček, M. and Whitelock, P., 2012, 'Setting up for corpus lexicography', *Proceedings of EURALEX 2012*, Oslo, Norway.
- Santos, D. and Bick, E., 2000, 'Providing Internet access to Portuguese corpora: the AC/DC project', *Proceedings of LREC 2000*, Athens, Greece.
- Verdelho, T., 2002, 'Dicionários portugueses, breve história', in J. H. Nunes and M. Petter (eds.) *História do saber lexical e constituição de um léxico brasileiro*, Humanitas.

Dictionaries

- Bluteau, R., 1712-1728, *Vocabulário português e latino, áulico, anatômico, architectonico*. Coimbra: Collegio das Artes da Companhia de Jesus. Digitized version available from <http://www.brasiliana.usp.br/en/dicionario/edicao/1>
- Borba, F. (ed.), 2004, *Dicionário UNESP do Português Contemporâneo*, 1st edition, Editora UNESP.
- Dicionário Escolar da Língua Portuguesa*, Academia Brasileira de Letras, 2009, Companhia Editora Nacional.
- Dicionário Priberam de Língua Portuguesa*, 2013, Available from: www.priberam.pt/dlpo/
- Ferreira Buarque de Hollanda, A. (ed.), 1975, *Novo Dicionário da Língua Portuguesa*, Nova Fronteira.
- Ferreira, M. and dos Anjos, M. (eds.), 2009, *Novo Dicionário Aurélio da Língua Portuguesa*, CD-ROM, Editora Positivo.
- Geiger, P. (ed.) 2012, *Novíssimo Aulete – Dicionário Contemporâneo da Língua Portuguesa*, Lexikon Editorial.
- iDicionário Aulete*, 2014, Lexikon Editorial, Available from: http://aulete.uol.com.br/site.php?mdl=aulete_digital
- Machado, J. (ed.), 1949-1959, *Grande Dicionário da Língua Portuguesa*, 12 volumes, 10th edition, Confluência.
- Malaca Casteleiro, J. (ed.), 2001, *Dicionário da Língua Portuguesa Contemporânea da Academia de Ciências de Lisboa*, Verbo.
- Morais Silva, A., 1789, *Diccionario da Lingua Portuguesa*. Digitized version available from <http://www.brasiliana.usp.br/en/dicionario/edicao/2>
- Novo Dicionário Lello da Língua Portuguesa*, 1996/1999, Lello Editores.
- Pequeno Dicionário Brasileiro da Língua Portuguesa*, 1938, Civilização Brasileira.
- Salles Villar, M. (ed.) *Dicionário Houaiss da Língua Portuguesa*, 2009, CD-ROM, Objetiva.
- Weiszflog, W. (ed.), 2004, *Michaelis Michaelis Moderno Dicionário da Língua Portuguesa*, Melhoramentos, Available from <http://michaelis.uol.com.br/>