A Morphological Historical Root Dictionary for Portuguese

João Paulo Silvestre, Alina Villalva
Centro de Linguística da Universidade de Lisboa
jpsilvestre@fl.ul.pt, alinavillalva@campus.ul.pt

Abstract

The project of a Morphological Historical Root Dictionary (MHRD) for Portuguese aims to build a specialized dictionary containing a critical selection of lexical units: the first stage is devoted to adjectives, namely those that can be found in Figueiredo (1913). Its main goals are the clarification of morphological and semantic issues in the evolution of the lexicon, and the assessment of the communicative adequacy of the words that are registered in current Portuguese dictionaries, particularly by signaling unused or seldom used words.

The methodology we established is three-sided: it first relies on the lexical analysis of the selected words; then, it seeks for lexicographic information in old Portuguese dictionaries (16th to 19th centuries), in Portuguese textual corpora and in etymological dictionaries; finally, it contrasts the Portuguese data with data from other romance languages, searching for lexical and semantic loans.

To conclude, we propose a prototype dictionary entry, applying the above-described methodology to the adjective *bravo.*

Keywords: word roots; morphology; Portuguese; historical lexicography

1   Preliminary remarks

The entry list of most contemporary Portuguese dictionaries still echoes lexicographic approaches that most of the other European languages have discarded along the past century. The main problem resides in the fact that they give room to the vast majority of the contents of previous dictionaries, thus accumulating, with an identical status, words that make part of the contemporary lexicon and a huge amount of unused words. This abundance of entries renders other problems usually found in dictionaries, such as graphic alternates, inadequate meanings and wrong etymological information, to name a few.

Furthermore, in the last two decades, mainstream dictionary publishing houses have rendered their workforce to the uncompelling issue of the orthographic 'entente' between Portugal, Brazil and (eventually) all other Portuguese-speaking countries, which motivated 'new', 'updated' paper editions. The remaining workforce of dictionary companies is fully devoted to the introduction of 'neologisms' that will haunt future editions.

There is an urge, then, to review the wordlists, to clean each entry, to correct the information given, to expunge old unused words. A dictionary may, of course, be cumulative about the selection of entries, but it must mark those that are unused, although they can be found in old literary texts.

2   Diachronic incoherence in Portuguese contemporary lexicography

Take, for instance, the case of the verb *abundar* 'to abound'. *Infopedia* is an online dictionary that claims to be the most complete dictionary of European Portuguese, covering general, technical and scientific vocabularies. Surprisingly, it registers words like *bondar, abondar* and *avondar* that are certainly not part of those vocabularies:

(1) Avondar. Verbo transitivo e intransitivo. Ver abundar. Do latim abundāre, «abundar» (*Infopédia*)

(2) Abondar. Verbo intransitivo. Regionalismo. Ser suficiente; bastar; bondar (Do latim abundāre, «idem» (*Infopédia*)

(3) Bondar. Verbo intransitivo. Popular. Bastar; ser suficiente (Do latim abundāre, «trasbordar; abundar» (*Infopédia*)

Probably, the form *avondar* is the oldest in Portuguese – it can be found in 14th to 17th century textual sources; *abondar* can also be found between the 14th century and the 19th; *abundar* starts in the 16th century and is the only form in contemporary usage[1]. Notice that the third form, i.e. *bondar*, marked in *Infopedia* as a 'popular' form has very few registers in the *Corpus do Português* database[2]. In fact, it has only two, one of which comes from an oral corpus, and it is quite difficult to understand.

A search in non-contemporary lexicographic sources hekps to consolidate the hypothesis above: *avondar* occurs in 16th and 17th century dictionaries, already marked as peripheral; *abondar* occurs in the 16th century, and in an 18th century it is considered as an error. The contemporary form, i.e. *abundar*, which is graphically closer to the spelling of the Latin verb (i.e. *abundare*), appears in the 17th century. Curiously, the recovery of all these variants began in 19th century dictionaries, such as Morais and Figueiredo. *Infopedia* replicates them, particularly Figueiredo.

(4) Auondar. Vide Abondar (Cardoso 1569)

(5) Avondar. Vide Abundar (Pereira 1697)

(6) Abundar. Ter abundancia. Erro, Abondar (Feijó 1734)

(7) Avondar, n. Abastar, ser bastante em numero. Antiq. [antiquate] (Silva 1831)

(8) Bondar v. i. Prov. Sêr bastante, sufficiente: mas isso não bonda. Alter. de abundar. (Figueiredo 1899)[3]

---

1   Examples were taken from the database *Corpus do Português.*
2   Only 7 matches found in texts, all from the 19th century (*Corpus do Português*).
3   Examples were taken from the database *Corpus Lexicográfico do Português.*

Filtering a general dictionary such as *Infopedia* is a desirable, but difficult task that requires highly trained manpower. This task is out of range for individual good will and low budget.

As we mentioned before, general contemporary Portuguese dictionaries either accumulate information from previous dictionaries, regardless of the errors they perpetuate and the real usage of the words, or they are produced on the basis of modern *corpora*, that integrate limited amounts of data, thus ignoring all the words that are not represented there. Lexical *corpora*, apart from coverage limitations, also frequently lack morphological tagging.

3 Planning a specialized historical dictionary

The MHRD project was designed in order to give a constructive response to such a negative perspective in the field of contemporary dictionary making. This project aims to build a specialized dictionary, which will contain a critical selection of the lexicon of Portuguese. Its main goals are:

- the clarification of the process of morphological and semantic evolution of the lexicon;
- the assessment of the communicative adequacy of the words that are registered in current Portuguese dictionaries, mostly by signalling unused or seldom used words.

MHRD will, thus, include simple and complex lexicalized roots, documented in a set of selected early lexicographic sources for Portuguese.

3.1 Lexicographic sources

Revisiting old dictionaries is thus obligatory, but instead of aiming to consider all of them, in the preliminary stage of this project, we decided to make a selection based on an analysis of lexicography in Portugal. The set of dictionaries that form this *canon* was selected for qualitative reasons, since they all played an important role either for the quality of the information they provided or for the normative role that they assumed. The selection, ranging from the 16th century to the end of the 19th century, includes:

- Jerónimo Cardoso (1569) *Dictionarium Latinolusitanicum / Lusitanicolatinum* — The first printed dictionary with extended word list in Portuguese (about 12 thousand entries). It is a testimony of ancient lexical choices and word forms, prior to the systematic imitation of Latin in neologisms and spelling.
- Bento Pereira (1697) *Prosodia in Vocabularium Bilingue, Latinum, et Lusitanum* — Extensive Latin-Portuguese learners dictionary, which represents the increase of lexical variety in the 17th century, by adapting many Latin words into Portuguese. In the corpus, there are about 50 thousand Portuguese word forms.

- Raphael Bluteau (1712-28) *Vocabulario Portuguez e Latino* — The first dictionary with examples from literary Portuguese texts, with more than 40 thousand entries. It makes a systematic collection of terminology and neologisms resulting from loans.

- António Morais Silva (1789) *Diccionario da Lingua Portugueza* — It is the first monolingual dictionary of the Portuguese, with a modern lexicographical technique. We also considered the fourth revised and extended edition of this dictionary, published in 1831. It was an authoritative dictionary throughout the nineteenth century. As a general rule, it notes old or unused words.

- Cândido de Figueiredo (1899) *Novo Diccionário da Língua Portuguesa* — A cumulative dictionary, which collects ancient and modern words without consistently noting their effective use in contemporary language. Served as lexical corpus to dictionaries in the 20th century, which reproduced the word list with little critical review.

Finally, in order to ascertain the usage of words in contemporary Portuguese, we consulted the *Corpus de Referência do Português Contemporâneo*. Completion and crosschecking of lexical analysis relies on the consultation of etymological dictionaries (Corominas and Pascual 1991) and historical dictionaries (*Le Trésor de la Langue Française Informatisé* and *El Nuevo Tesoro Lexicográfico de la Lengua Española*).

3.2 Root identification

Since feasibility is one of our main concerns, we decided to limit our research to simple (unanalysable) roots. We believe that, once we have achieved to isolate the core set of roots, we will be better equipped to identify and describe derived and compound words. Simple words, those that are projected from single roots, have a supplementary advantage: they are usually old words and old words tend to accumulate or to change meanings. Those moments, which are not easy to detect, are seldom documented.

The identification of the core set of roots is certainly crucial for a better understanding of the Portuguese lexicon, but no existing general or specialised Portuguese dictionary or lexical corpus provides this information. Two specialized dictionaries deserve to be mentioned, however. The first one is the *Dicionário de Raízes e Cognatos* [cf. Goes (1921)]. Apart being based on 19th century sources, it mainly deals with the small subset of neoclassical roots. The second one is the most important Portuguese morphological dictionary. It was made in Brazil, by gathering lexical information from unspecialized sources, such as general language dictionaries [cf. Heckler, Back, Massing (1984-1988)]. Both of them offer interesting data, but their consultation also needs extensive critical reading.

Most Portuguese words, irrespectively of their longer or shorter existence in the Portuguese lexicon, come from a relatively stable set of roots, which can be documented in morphologically simple words (as free roots) or in complex words (as free roots in compositional words and as bound roots in lexicalized words). This typology of roots (based on Villalva and Silvestre [in print]) also foresees cases of bound roots in compositional complex words:

Root	Head		Complement		Modifier	
	Free root	Bound root	Of a derivational suffix	Of another root	Of a simple word	Of a complex word
Type 1 e.g. rat-	rat-o 'mouse'		rat-ice 'cunning'	rat-icida 'rat poison'		
Type 2 e.g. gastr-			gástr-ico 'gastric'	gastr-onomia 'gastronomy'		
Type 3 e.g. graf-	graf-ar 'to write'	pluvió-graf-o 'rain gauge recorder'	gráf-ico 'graphic'	graf-ologia 'graphology'		
Type 4 e.g. super-					super-amigo 'super-friend'	super-confortável 'super-comfortable'

Table 1: Typology of roots.

Therefore, a specific methodology had to be established. Bearing feasibility in mind, we decided to devote our initial research to adjectives. The approach we decided to take is three-sided: it relies on the lexical analysis of the selected words; it seeks lexicographic information in old Portuguese dictionaries (16th to 19th centuries) and Portuguese textual *corpora* and in etymological dictionaries; it contrasts the Portuguese data with data from other romance languages, searching for lexical and semantic loans.

The first stage of the project is devoted to adjectives, the second to nouns and the third one to verbs.

3.3 Adjectives

The lexical analysis of adjectives considers their morphological, syntactic and semantic properties. Notice that, from a morphological point of view, Portuguese adjectives are not significantly different from nouns, which raises a practical problem for the selection of roots.

Adjectives and nouns, they both require number inflection (cf. Table 2, i) and they both comprise a subset that allows for gender variation (cf. ii) and a subset of invariable forms (cf. iii).

i)	ii)	iii)
casa $_{\text{Nsingular}}$ 'house' casas $_{\text{Nplural}}$ 'houses' leve $_{\text{ADJsingular}}$ 'light' leves $_{\text{ADJplural}}$ 'light'	gato $_{\text{Nmasculine}}$ 'male cat' gata $_{\text{Nfeminine}}$ 'female cat' novo $_{\text{ADJmasculine}}$ 'new' nova $_{\text{ADJfeminine}}$ 'new'	casa $_{\text{Nfeminine}}$ 'house' carro $_{\text{Nmasculine}}$ 'car' leve $_{\text{ADJ}}$ 'light'

Table 2: Adjectives - morphology.

There are, nevertheless, differences that can be spotted. As far as number is concerned, although its specification is compulsory for nouns and adjectives alike, in nouns it has semantic relevance (singular refers one entity, plural refers more than one), in adjectives it is semantically irrelevant: adjectives have no quantifiable meaning - number inflection is merely relevant for syntactic agreement ($casa_{singular}$ $nova_{singular}$ 'new house'; $casas_{plural}$ $novas_{plural}$ 'new houses').

Gender is even more diverse. All nouns have to have a gender value, which is lexically determined, irrespective from their possibility to participate in gender contrasts (cf. *gato* / *gata*; *carro* and *casa*, in table 3, i). In general, animate nouns can participate in gender contrasts either by thematic alternation (cf. i), by a morphological resource (cf. ii) or lexically (cf. iii), but some animate nouns do not, which eventually creates a mismatch between grammatical gender and the gender of the referent (cf. iv):

i)	ii)	iii)	iv)
gato'male cat' gata 'female cat' aluno 'male student' aluna 'female student'	galo 'rooster' galinha 'hen' marquês 'marquis' marquesa 'marchioness'	cavalo 'horse' égua 'mare' homem 'man' mulher 'woman'	testemunha $_{feminine}$ 'witness (male or female) cônjuge $_{masculine}$ 'spouse (husband or wife) águia $_{feminine}$ 'eagle (male or female)' águia-macho $_{feminine}$ 'male eagle' águia-fêmea $_{feminine}$ 'female eagle' rinoceronte $_{masculine}$ 'rhino' rinoceronte-macho $_{masculine}$ 'male rhino' rinocerente-fêmea $_{masculine}$ 'female rhino'

Table 3: Adjectives – gender contrast.

Inanimate nouns are never allowed to participate in gender contrasts. It is possible to find pairs of words that apparently share the same root, although they belong to different thematic classes. They are not in a gender contrast - they are different words ($casa_{feminine}$ 'house'; $caso_{masculine}$ 'case').

For adjectives, gender is as irrelevant as number. It may be syntactically important, for agreement, but a great deal of adjectives is invariable, so the syntactic relevance is also questionable – it is probably just a vestige from Latin declension.

Apart from these morphosyntactic properties, adjectives and nouns also share the possibility to undergo evaluative affixation. It is particularly relevant to notice that the most productive suffix (i.e. *–inh{o, a}(s)* is equally available, but its semantic effect on nouns differs from its semantic outcome in adjectives. In the first case, it is typically a diminutive or valuative (cf. table 4, i); in the second case its reading is ambiguous – typically, it can either be an attenuative or a superlative (cf. Table 4, ii). On the other hand, the superlative forming suffix (i.e. *-íssimo*) only adjoins to adjective bases, but a large set of unquestionable adjectives are not scalable and thus they do not allow the adjunction of this suffix (cf. Table 4, iii).

i)	ii)	iii)
casinha 'small house' carrinho 'small car' gatinho 'small (dear) cat'	novinho 'pretty/very young'	*casíssima 'very+house' novíssimo 'very+new' *teatralíssimo 'very+theatrical'

Table 4: Adjectives - afixation.

In sum, morphology can provide some clues to help setting adjectives apart from nouns, but it fails to draw a neat borderline. Syntactic distribution has, thus, to be considered as well, in order to characterise adjectives as a word class; it is also relevant to establish adjective subclasses. The next set of examples comprises a subset of words that can only occur in an adjective position and another subset of nouns that never occur as adjectives (*cabelo*$_N$ *fino*$_{ADJ}$ 'thin hair'); a third subset includes words that occur in adjectival and nominal contexts (*professor*$_N$ *assistente*$_{ADJ}$ 'assistant professor'; *assistente*$_N$ *do produtor*$_N$ 'producer's assistant'):

In order to get a better understanding of adjectives on the basis of syntactic criteria, probably the most relevant are those that concern word order and the possibility to occur in predicative positions as well as in non-predicative positions. Colour adjectives, for instance, can never occur in a prenominal position (*vestido vermelho* 'dress red'; **vermelho vestido* 'dress red'), but other adjectives can (*velho hábito* 'old habit'), although this is a marked word order, except for ordinal adjectives (*primeiro dia* 'first day'; **dia quinto* 'fifth day'):

The predicative *vs.* non-predicative distinction also fails to clearly set adjective subclasses: most adjectives can have both distributions (*o vestido vermelho* 'the red dress'; *o vestido é vermelho* 'the dress is red') and small sets of adjectives have exclusive distribution (*o primeiro dia* 'the first day; *o vestido é vermelho* 'the dress is red').

Finally, we need to consider the semantics of adjectives, which is probably the most difficult aspect to deal with. Several ontologies have been suggested in the literature, but none of them is able to avoid very specific world knowledge constraints.

We will also use a set of criteria (somehow in parallel with morphological and syntactic criteria above considered) that will apply to each form. The first condition concerns gradability, measured on the basis of *–íssimo* affixation and also on the basis of syntactic comparative constructions. This condition allows us to identify three subsets of adjectives: those that respond positively to both tests (cf. table 5, i), those that respond positively just to *–íssimo* (cf. ii)[4] and those that respond negatively to both of them (cf. iii). Notice that this condition has to be tested in a specific syntactic context, since the result is not always the same (cf. iv):

---

4   This contrast is probably due to the fact that the evaluative suffix is closer to a rhetoric resource than to the setting of a degree.

i)	preço alto 'high price' preço altíssimo 'very+high price' preço muito alto 'very high price'
ii)	casa enorme 'enormous house' casa enormíssima 'very+enormous house *casa muito enorme 'very enormous house'
iii)	sais minerais 'mineral salts' *sais mineralíssimos 'very+mineral salts' *sais muito minerais 'very mineral salts'
iv)	os cavalos estão cansados 'the horses are tired' os cavalos cansados não ganham corridas 'tired horses don't win races' os cavalos estão cansadíssimos 'the horses are very+tired' *os cavalos cansadíssimos não ganham corridas 'very+tired horses don't win races'

Table 5: Adjectives – *íssimo* afixation.

The second condition concerns the possibility to relate an adjective to another adjective, usually by opposition. Rasken and Nirenburg distinguish binary oppositions of non-gradable, complementary antonyms (cf. table 6, i), polar oppositions of gradable antonyms (cf. ii) and multiple non-gradable oppositions (cf. iii):

i)	ii)	iii)
morto vs. vivo 'dead' vs 'alive'	(*muito frio*) *frio* (*pouco frio*) vs. (*pouco quente*) *quente* (*muito quente*) 'very cold' 'cold' 'bit cold' vs. 'bit hot' 'hot' 'very hot'	*análise histórica* 'historical analysis' .... *económica* 'economic' ... . *social* 'social' ... . *política* 'political' ...

Table 6: Adjectives – sense relations.

Finally, we need to consider the semantic features of the antecedent, since this information is of crucial importance to circumscribe the meaning of adjectives. In particular, it is necessary to identify the value of animacy, humanness, countability, concreteness. Notice, for instance, that the adjective *bravo* (described in some detail in Villalva & Silvestre 2011) has accumulated different meanings, ranging from a negative pole to a positive pole. In the first case, it means 'ferocious' if it applies to animals (wild animals), but it will mean 'angry' when it gets to be applied to people.

(9) $N_{[-human]}$ *bravo* 'ferocious' (Cardoso 1569)

(10) $N_{[+human]}$ *bravo* 'courageous' (Bluteau 1712-1728)

(11) $N_{[+human]}$ *bravo* 'angry' (Silva 1789)

3.4 *–mente* adverbs

The classification of deadjectival adverbs may also bring some problems. Consider the case of *alto*, which is a good representative of a set of adjectives that are used to measure dimension, in its various features (height, length, weight, etc.). This is a word of Latin origin (i.e. *altus*), originally a participle of the verb *alo* 'to feed', thus meaning 'fed, grown'. As an adjective, *altus* had two basic meanings[5], related to the perception of height (A. Seen from below upwards, *high* and B. Seen from above downwards, *deep*). Both of them allowed a figurative, non-physical interpretation, respectively 'elevated, distinguished' and 'profound'.

The semantic interpretation of the Portuguese adjective *alto* (as well as the Spanish cognate, *alto*, the French *haut* or the Italian *alto*) replicates the semantics of the A. interpretation of the Latin *altus*. Derived nouns, such as *altura* 'height' and *alteza* 'highness', as well as derived verbs such as *altear* 'to heighten' and *enaltecer* 'to praise', help to consolidate that conclusion.

If no other reason existed, the adjective *alto* would not deserve much more comment, but that is not the case if we also consider the derived adverb *altamente*. All romance languages share an adverb forming resource which is based in the grammaticalization of a Latin noun (*mens, mentis*) that took place prior to the individuation of Romance languages, thus explaining its vitality still in contemporary strata. In Latin, the ADJ + *mente* sequence had an adverbial usage. Probably, *-mente* adverbs (or adverbs to be) in Romance languages were initially strictly manner adverbs. Contemporary usage is a bit more complex: apart from manner adverbs (e.g. *elegantemente* 'elegantly', *injustamente* 'unfairly'), *-mente* also forms temporal locatives such as those in table 7, generally equivalent to a locution that includes the base adjective (i.e. *anteriormente* = num momento anterior; *imediamente* = num momento imediato; *futuramente* = num momento futuro).

Past	Present	Future
anteriormente	actualmente	brevemente
antigamente	presentemente	futuramente
inicialmente		seguidamente

Table 7: Adverbs - Temporal locatives.

Although the formation of –mente adverbs is considered as a very productive word formation process, some restrictions have been identified. According to Scalise (1990), Italian *-mente* cannot be attached to several types of adjectives. Unsurprisingly, the same negative constraints apply in Portuguese, so *–mente* cannot apply to possessive (cf. **minhamente*), demonstrative (cf. **estamente*), indefinite (cf. **qualquermente*), or numeral adjectives (cf. **umamente*), nor can it be attached to qualifying adjectives denot-

---

5   http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.04.0059%3Aentry%3Daltus1

ing physical qualities (cf. *gordamente*), colour adjectives (cf. *vermelhamente*), superlatives (cf. *melhormente*) and adjectives modified by evaluative suffixes (cf. *ligeirinhamente*).

Considering these negative constraints, the adjective alto should not yield a –*mente* adverb, but *altamente* is a quite frequently used word. Scalise (1990) considers that polysemic adjectives that refer a physical property and a psychological property can derive a –*mente* adverb from the second meaning[6]. In fact, *altamente* is never related to the physical meaning of *alto* – it never means 'in a high manner'; but it not related to the psychological meaning of *alto* – it never means 'in an elevated manner'.

Evidences from ancient dictionaries confirm that there are only occurrences of polysemous meanings of *altamente* and the same applies to other adjectives indicating size or extension, such as *baixo, largo, estreito, comprido, longo, curto.* Should be noted that in the same sources the adverbs *largamente* and *longamente* are considered as synonyms (cf. 21).

(12) Profundè. aduer*. Alta* & fundamente. (Cardoso 1569)

(13) Tragice loqui Falar *altamente.* como em tragedia. (Cardoso 1569)

(14) Eminenter, adv. Excellentemente, *altamente.* (Pereira 1697)

(15) *Altamente.* Alte. Sublimiter. (Pereira 1697)

(16) *Baixamente.* Abjecte. Ignobiliter. (Pereira 1697)

(17) Spatiose, Adv. *Larga*, espaçozamente. (Pereira 1697)

(18) Largiter, Adv. *Larga*, liberal, *abundantemente.* (Pereira 1697)

(19) *Estreitamente.* Anguste. arcte. (Cardoso 1569)

(20) Conjunctissime, adv. superl. Muito junta, & *estreitamente*, muito amigavelmente. (Pereira 1697)

(21) Prolixe, Adv. *Larga,   longa, liberalmente.* (Pereira 1697)

(22) Pleniter. Adv. Plena, cheia, copioza, perfeita, *compridamente.* (Pereira 1697)

(23) Perlonge, Adv*.* Mui longa, & *compridamente*; muito longe. (Pereira 1697)

(24) *Curtamente.* Timide. (Pereira 1697)

CRPC returns more than 8.000 matches for *altamente* and two conclusions become self-evident: it is always a modifier of an adjectiv,e and it is always a quantifier adverb, equivalent to *muito* 'very'.

(25) *situação* **altamente** *benéfica* 'highly beneficial situation' (*CRPC*)

(26) *factores* **altamente** *estimulantes* 'highly stimulating factors'(*CRPC*)

In fact, most –*mente* adverbs that occur as adjective modifiers has exactly the same meaning as *altamente,* which demonstrates that the meaning of the adverb is not related to the meaning of the base adjective:

(27) *a praia continua* **abençoadamente** *vazia* 'the beach continues blessedly empty' (*CRPC*)

(28) *convite tão* **abertamente** *amoroso* 'such openly loving invitation' (*CRPC*)

(29) *coisas* **abissalmente** *diferentes* 'abysmally different things' (*CRPC*)

---

6   Scalise (1990) presents the eaemple of aridamente that is related to the meaning 'boring' and not to the meaning 'dry' of the adjective arido.

3.5   Entry structure: the case of *bravo*

The structure of articles aims to gather information about the roots and derived words, revealing the diachronic sequence and semantic relations. The lexical research that precedes the compilation of an article can be exemplified in previous works about the semantic evolution of the adjective *bravo* (Villalva, Silvestre 2011) and about adjectives that have undergone a severe meaning shift, like *esquisito* (Silvestre, Villalva, in print). In this paper, we propose a prototype entry, applying the data collected on *bravo* (see table 8).

The entry headword is the root. The first category is the information on the simple word, indicating word class, etymology and a summary of documented semantic evolution. Complex words formed directly from the base (such as -mente adverbs) have a separate description. The other groups are the complex adjectives, verbs and names. For each of the words identified, we provide the date of first attestation in the selected dictionaries, as well as the semantic equivalence. Based on the occurrence in lexical corpora (especially CRPC), we finally evaluated the word frequency. The result is an assessment on the contemporary use of words (unused forms are explicitly marked), with which we intend to contribute to the review of dictionary wordlists and to improve word formation descriptions.
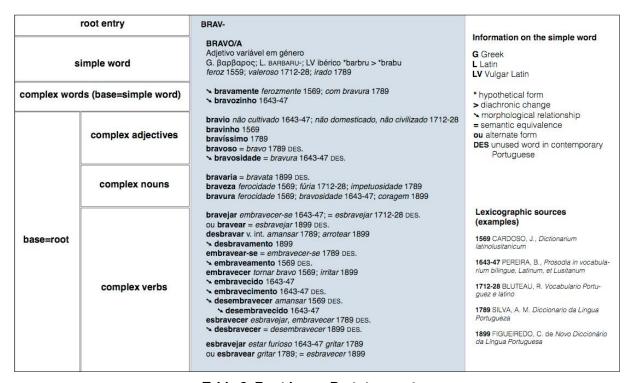
root entry		BRAV-	Information on the simple word
simple word		**BRAVO/A** Adjetivo variável em gênero G. βαρβαρος; L. BARBARU-; LV ibérico *barbru > *brabu *feroz* 1559; *valeroso* 1712-28; *irado* 1789	**G** Greek **L** Latin **LV** Vulgar Latin
complex words (base=simple word)		↘ **bravamente** *ferozmente* 1569; *com bravura* 1789 ↘ **bravozinho** 1643-47	∗ hypothetical form > diachronic change ↘ morphological relationship = semantic equivalence **ou** alternate form **DES** unused word in contemporary Portuguese
base=root	complex adjectives	**bravio** *não cultivado* 1643-47; *não domesticado, não civilizado* 1712-28 **bravinho** 1569 **bravíssimo** 1789 **bravoso** = *bravo* 1789 DES. ↘ **bravosidade** = *bravura* 1643-47 DES.	
	complex nouns	**bravaria** = *bravata* 1899 DES. **braveza** *ferocidade* 1569; *fúria* 1712-28; *impetuosidade* 1789 **bravura** *ferocidade* 1569; *bravosidade* 1643-47; *coragem* 1899	**Lexicographic sources (examples)**
	complex verbs	**bravejar** *embravecer-se* 1643-47; = *esbravejar* 1712-28 DES. ou **bravear** = *esbravejar* 1899 DES. **desbravar** v. int. *amansar* 1789; *arrotear* 1899 ↘ **desbravamento** 1899 **embravear-se** = *embravecer-se* 1789 DES. ↘ **embraveamento** 1569 DES. **embravecer** *tornar bravo* 1569; *irritar* 1899 ↘ **embravecido** 1643-47 ↘ **embravecimento** 1643-47 DES. ↘ **desembravecer** *amansar* 1569 DES. ↘ **desembravecido** 1643-47 **esbravecer** *esbravejar, embravecer* 1789 DES. ↘ **desbravecer** = *desembravecer* 1899 DES. **esbravejar** *estar furioso* 1643-47 *gritar* 1789 ou **esbravear** *gritar* 1789; = *esbravecer* 1899	**1569** CARDOSO, J., *Dictionarium latinolusitanicum* **1643-47** PEREIRA, B., *Prosodia in vocabularium bilingue, Latinum, et Lusitanum* **1712-28** BLUTEAU, R. *Vocabulario Portuguez e latino* **1789** SILVA, A. M. *Diccionario da Lingua Portugueza* **1899** FIGUEIREDO, C. de *Novo Diccionário da Língua Portuguesa*

Table 8: Root brav - Prototype entry.

4 References

Bluteau, R. (1712-1728). *Vocabulario portuguez e latino*. Coimbra-Lisboa, Collegio das Artes da Companhia de Jesu

Cardoso, J. (1569-1570). Dictionarium latinolusitanicum & vice versa lusitanico latinum. Conimbricae, Joan. Barrerius

Corominas, J., Pascual, J. A. (1981). *Diccionario crítico etimológico castellano e histórico*. Madrid, Gredos.

Corpus de Referência do Português Contemporâneo. Accessed at: http://www.clul.ul.pt/pt/recursos/183-crpc#cqp

Corpus do português: 45 million words, 1300s-1900s. Accessed at: http://www.corpusdoportugues.org [10/03/2014].

Corpus Lexicográfico do Português. Accessed at: http://clp.dlc.ua.pt [10/03/2014].

Figueiredo, C. de (1913). *Novo diccionário 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.

Infopédia. Dicionário da Língua Portuguesa da Porto Editora. Accessed at: http://www.infopedia.pt/lingua-portuguesa/ [10/03/2014].

[10/03/2014].

Le Trésor de la langue française informatisé. Accessed at: http://atilf.atilf.fr/tlf.htm [10/03/2014].

Nuevo tesoro lexicográfico de la lengua española. Accessed at: http://ntlle.rae.es/ntlle/SrvltGUILoginNtlle [10/03/2014].

Pereira, B. (1697). Prosodia in vocabularium bilingue, latinum et lusitanum. Eborae, Typographia Academiae

Raskin, V., Nirenburg, S. (1995). *Lexical Semantics of Adjectives: A Microtheory Of Adjectival Meaning*. Memoranda in Computer and Cognitive Science MCCS-95-288. Las Cruces - New Mexico, New Mexico State University.

Scalise, S. (1990) Constraints on the Italian suffix –mente. In W. U. Dressler (ed.), *Contemporary Morphology*. Berlim, Mouton de Gruyter.

Silva, A. M. (1789). *Diccionario da lingua portugueza.* Lisboa, na Of. de Simão Thaddeo Ferreira

Silvestre, J. P., Villalva, A. (in print). Mutations lexicales romanes: esquisito, bizarro et comprido. In *InnTrans: Innsbrucker Beiträge zu Sprache, Kultur und Translation*. Peter Lang Verlag.

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. In *ReVEL* v. 9, n. 17. Accessed at: http://www.revel.inf.br/files/artigos/revel_17_de_bravo_a_brabo.pdf [10/03/2014].

Villalva, A., Silvestre, J. P. (in print). *Introdução ao estudo do léxico. Descrição e análise do Português.* Petrópolis-Rio de Janeiro, Vozes.