

10 The Phonology and Morphology of Word Formation

ALINA VILLALVA AND CARLOS
ALEXANDRE GONÇALVES

Portuguese shares many morphological features with other Romance languages, such as Castilian, Italian, or French, but it also displays properties that set it apart from the other members of the Romance family. In this chapter, we will privilege the latter aspects. The resemblance with other romance languages and, at the same time, the specificity of Portuguese morphology further echoes in the comparison of different national varieties, such as the European (henceforth EP) and Brazilian Portuguese (henceforth BP) subsystems.

Affixation and compounding are the main word-formation processes in Portuguese.¹ We will present an overview of their main morphological and phonological properties and also some border issues, indicating, when appropriate, contrasts between EP and BP (with reference to the dialects of Lisbon and southern Brazilian variants). In addition, we will discuss some types of word formation not addressed by the grammatical tradition. While influential traditional studies such as Mattoso Câmara (1971) and Basílio (1987), for BP, or Carvalho (1967) and Rio-Torto (1998), for EP, are discussed where appropriate, the discussion in this chapter is especially based on Villalva (1994), Gonçalves (2004), Gonçalves (2012), and Villalva and Silvestre (2014).

In our exposition of the Portuguese word-formation processes, we assume that words (W), as morphological structures, are projections of the root (R), which is morphologically specified by a thematic constituent² (TC) that generates a stem (S). The stem is then morpho-syntactically (MSS)specified.³ This is the underlying morphological structure of all simple words:

$$(1) \quad [[[X]_R [Y]_{TC}]_S]_{MSS}]_W$$

Roots are lexical units, specified to a large number of features (their phonological representation and morphological, syntactic and semantic features, among others). One of these features concerns the thematic class to which they belong. Verbs are assigned to a conjugation class (first, second or third), a distinction that has no syntactic or semantic consequences—it is relevant merely for inflection (the phonetics are of the EP variant):

(2)	1st C	2nd C	3rd C
infinitive	<i>can't[a]r</i> "to sing"	<i>be'b[e]r</i> "to drink"	<i>fu'g[i]r</i> "to run away"
pres.ind. 3rds.	'cant[v]	'beb[i]	'fog[i]
pres.subj.1st/3rd s.	'cant[i]	'beb[v]	'fu][v]
past participle	<i>can't[a]do</i>	<i>be'b[i]do</i>	<i>fu'g[i]do</i>

Although this lexically determined distinction is not much more than a remnant of the Latin conjugation system, derivation is still sensitive to it. Deverbal suffixes that select verb stems specify a subclass of stems: infinitive stem (IST), present stem (PRST) and past stem (PTST). First conjugation verbs neutralize this distinction, but the thematic vowel of second and third conjugation inflected forms has different phonetic outputs:

(3)	1st C	2nd C	3rd C
IST	<i>apresen't[ar]</i> "to present"	<i>absol'v[er]</i> "to absolve"	<i>defi'n[i]r</i> "to define"
PRST	<i>apre'sent[v]</i> "it presents"	<i>ab'solv[i]</i> "it absolves"	<i>de'fin[i]</i> "it defines"
PTST	<i>apresen't[ado]</i> "presented"	<i>absol'v[i]do</i> "absolved"	<i>defi'n[i]do</i> "defined"

Deverbal suffixes are sensitive to this subclass feature:

(4)	1st C	2nd C	3rd C
IST	<i>ado'ç[ente]</i> "sweetener"	<i>reque'r[ente]</i> "petitioner"	<i>pe'd[i]nte</i> "beggar"
PRST	<i>igno'r[encia]</i> "ignorance"	<i>inci'd[encia]</i> "incidence"	<i>emer'g[encia]</i> "emergency"
PTST	<i>dedic[v]ç[ão]</i> "dedication"	<i>absolv[i]ç[ão]</i> "absolving"	<i>defin[i]ç[ão]</i> "definition"

Nouns and adjectives split over a larger number of classes, also lexically determined, which are related to the nature of the thematic vowel and to the gender value.⁴ Even though final *-o* ([u]) and final *-a* ([v]) have long been identified as gender morphemes, it is easy to conclude that they do not have that status: *-o* ending words are typically masculine, but masculine nouns may have many other endings. The same occurs with *-a* ending words and feminine. In the following table, all possibilities are registered for nouns:

(5)	Masculine	Feminine	Non-Specified
<i>-o</i> ([u])	<i>ca'val[u](s)</i> "horse(s)"	' <i>trib[u](s)</i> "tribe(s)"	<i>sol'dad[u](s)</i> "soldier(s)"
<i>-a</i> ([v])	<i>pro'fet[v](s)</i> "prophet(s)"	' <i>vac[v](s)</i> "cow(s)"	<i>a'tlet[v](s)</i> "athlete(s)"
<i>-e</i> ([i, I])	EP: ' <i>pent[i](s)</i> BP: ' <i>pent[I](s)</i> "comb(s)"	EP: ' <i>pel[i](s)</i> BP: ' <i>pel[I](s)</i> "skin(s)"	EP: <i>a'gent[i](s)</i> BP: <i>a'gent[I](s)</i> "agent(s)"
\emptyset ([], [i, I])	EP: <i>tra'tor[]([i]s)</i> BP: <i>trato'r[]([I]s)</i> "tractor(s)"	' <i>paz[]([i]s)</i> ' <i>paz[]([I]s)</i> "peace(s)"	<i>fis'ca[I]([j]s)</i> <i>fis'ca[w]([l]s)</i> "supervisor(s)"
athematic (\mp)	<i>'pau(s)</i> "stick(s)"	' <i>pá(s)</i> "shovel(s)"	<i>refém(s)</i> "hostage(s)"

The difference between *-o* stems and *-a* stems is quite obvious: there is a final [u] and a final [v], respectively in the words where they are present. The class of *-e* stems features the thematic index [i] in EP, [I] in BP. The class of \emptyset stems features a thematic index that triggers a high vowel ([i] in EP, [I] in BP) with no phonetic realization in final position (singular), except for [l] ending roots. These two classes are very similar and they might be considered as one if all [l], [r] or [s] ending roots were \emptyset stems, but this is not the case ('*pele*/*pa'pel* "skin/paper"; *fol'clore*/*'flor* "folklore/flower"; '*gás*/*'gaze* "gas/gauze"). In EP, the endings of words such as '*pele* and *pa'pel* are phonetically very similar in the singular ([pɛl]/[pɛ'pɛl]), but they differ in the plural ([pɛli]/[pɛ'pɛj]). The difference is easier to understand if these two words are assigned to different thematic classes. If their underlying representation is /pɛl+i/ and /pɛpɛl+/, then the plural of /pɛpɛl+/ can be obtained by semi-vocalization of the final consonant of the root, but only if it is a \emptyset root. Notice that orthography is irrelevant: the majority of *-e* roots ends in a graphic <e>, but words such as *aval* (pl. *avales*) "approval" or *fel* (pl. *feles*) "gall" are also *-e* stems. Athematic stems are easier to recognize: they have neither a thematic index nor a trace of one.

The status of thematic classes is quite peculiar, since they have no syntactic or semantic relevance, and, from a phonological point of view, thematic indices (i.e. *-a*, *-o*, *-e*) are

uninteresting elements, since they are always unstressed vowels that show up at the right border of the word. Yet, thematic classes are morphologically relevant, both for inflection⁵ and for word formation. As we will see below, the choice of roots, stems (including subtypes of stems for deverbal derivation) or words is part of the selectional constraints of affixes.

Complex words expand the structure in (1). In Portuguese, most word-formation processes occur in the root domain, and they may involve a root and an affix or several roots. Processes involving a single root usually attach an affix, which can be a morphological predictor (commonly known as derivational suffix) or a morphological modifier (a prefix or an evaluative suffix). Those involving more than one root are morphological compounding processes. This structural distinction requires a neat demarcation of roots and affixes. It also requires the identification of the grammatical roles for word constituents.

1. Affixation

Affixation is traditionally described as involving suffixation (which is predominant in Portuguese), or prefixation,⁶ but this topological description needs to be complemented by a grammatical analysis. In fact, word-formation affixes can be predictors, which means that they are the head of the structure they generate, or modifiers. In Portuguese, all predictors are suffixes (= derivational suffixes), all prefixes are modifiers, and some suffixes (= evaluative suffixes) are modifiers.

1.1. Derivational suffixation

In Portuguese, derived words are generated on the basis of derivational suffixes according to their selectional and inherent properties. Selectional properties are the set of constraints involving the base form. Base forms can be roots (adjective roots, as in (6a); noun roots, as in (6b); verb roots, as in (6c)), stems (only verb stems are available, as in (6d–f)) or words (just adjectives, as in (6g)). Inherent properties define the features of the output, which is always a root that will be projected into a stem, first, and then to a word: derived forms can be adjective, adverb, noun, or verb roots (stressed syllables are in bold):

- | | | | |
|--------|---|---|--------------------------|
| (6) a. | [<i>humanist</i>] _{ADJR} → | [[[<i>huma'nístic</i>] _{ADJR} o] _{ADJS}] _{ADJ} | “humanist→humanistic” |
| | [<i>ingenu</i>] _{ADJR} → | [[[<i>ingenui'dad</i>] _{NR} e] _{NS}] _N | “naïve→naiveté” |
| | [<i>fragil</i>] _{ADJR} → | [[[<i>fragili'z</i>] _{VR} a] _{VS}] _V | “fragile→to weaken” |
| b. | [<i>gost</i>] _{NR} → | [[[<i>gos'tos</i>] _{ADJR} o] _{ADJS}] _{ADJ} | “taste→tasty” |
| | [<i>arroz</i>] _{NR} → | [[[<i>arro'zal</i>] _{NR}] _{NS}] _N | “rice→rice field” |
| | [<i>frut</i>] _{NR} → | [[[<i>frutifi'c</i>] _{VR} a] _{VS}] _V | “fruit→to fructify” |
| c. | [<i>mand</i>] _{VR} → | [[[<i>man'dão</i>] _{NR}] _{NS}] _N / [[<i>man'don</i>] _{NR}] _a] _{NS}] _N | “to boss→bossy” |
| d. | [<i>dança</i>] _{VSINF} → | [[[<i>dan'çânt</i>] _{ADJR} e] _{ADJS}] _{ADJ} | “to dance→dancing” |
| | [<i>grava</i>] _{VSINF} → | [[[<i>grava'dor</i>] _{NR}] _{NS}] _N | “to engrave→engraver” |
| e. | [<i>concorda</i>] _{VSPRES} → | [[[<i>concor'dânci</i>] _{NR}] _a] _{NS}] _N | “to agree→agreement” |
| f. | [<i>procura</i>] _{VSPAST} → | [[[<i>procu'rável</i>] _{ADJR}] _{ADJS}] _{ADJ} | “to search→searchable” |
| | [<i>separa</i>] _{VSPAST} → | [[[<i>separa'ção</i>] _{NR}] _{NS}] _N | “to separate→separation” |
| g. | [<i>urgente</i>] _{ADJ} → | [[[<i>ur'gente'ment</i>] _{ADVR} e] _{ADVS}] _{ADV} | “urgent→urgently” |

Selectional properties vary from suffix to suffix. They can make use of, at least, phonological/prosodic properties of the base (see below the allomorphy of *-ez* ~ *-eza*, for instance), syntactic properties (especially for deverbal suffixes that require information on the argument structure of the base verb) and semantic properties (the collective noun forming suffix *-agem* selects the root of count nouns: *'folha*→*fo'lhagem* (“leaf→foliage”).

Besides defining the syntactic category of the output, derivational suffixes also participate in broad semantic categories. The existence of competing suffixes occurs inside these categories:

(7) causative verbs:

- [[*escur*]_{ADJR} *e*'*c*]_{VR} *er*_V "to darken"
 [[*agil*]_{ADJR} *i*'*z*]_{VR} *ar*_V "to hasten"
 [[*solid*]_{ADJR} *ifi*'*c*]_{VR} *ar*_V "to solidify"

action nouns:

- [[*apresenta*]_{VS} *'çãõ*]_{NR} "presentation"
 [[*esqueci*]_{VS} *'ment*]_{NR} *o* "forgetfulness"
 [[*tole'ra*]_{VS} *'nci*]_{NR} *a* "tolerance"

subject nouns:

- [[*apresenta*]_{VS} *'dor*]_{NR} "presenter"
 [[*representa*]_{VS} *'nta*]_{NR} *e* "representative"

Derivational processes available in EP and BP are virtually identical: they share the same set of suffixes, and their behavior is quite similar. There is, however, a margin of contrasts that is worth noting. It is quite common to find different suffixes competing within the same morphosemantic category:

- (8) EP: *dese*'*nhista* / BP: *dese*'*nhista* "designer"
 EP: *fuma*'*dor* / BP: *fu*'*mante* "smoker"

Thus, derived words in EP and BP are derived autonomously, yielding different results within a given morphosemantic category. Another distinction is set by the mutation of some affixes—for instance, in BP, the suffix *-d(a)*, in the expression *X-da* ((*dar uma*) *olhada* "to take a quick look"), forms brief action nouns. No such suffix exists in EP, although there is a semantically equivalent suffix, which is *-dela* ((*dar uma*) *olhadela*), not used in BP.

1.2. Parasynthesis

Parasynthesis is a particular case of derivation, usually defined as a process of simultaneous prefixation and suffixation. However, considering that sometimes no suffix intervenes and that the prefix is an expletive element, this type of derivation shows a striking resemblance with suffixation or conversion, except for the fact that it requires the presence of the expletive prefix:

- (9) a. [prefix[[ADJR]suffix]_{VR}]_{VR}
 [[*a*[[*mol*]_{ADJR} *e*'*c*]_{VR}]_{VR} [*e*]_{VS} [*r*]_V "to soften"
 [[*en*[[*rouqu*]_{ADJR} *e*'*c*]_{VR}]_{VR} [*e*]_{VS} [*r*]_V "to hoarsen"
 [prefix[[ADJR]]_{VR}]_{VR}
 [[*a*[[*cele*'*r*]_{ADJR}]_{VR}]_{VR} [*a*]_{VS} [*r*]_V "to accelerate"
 [[*en*[[*ri*'*j*]_{ADJR}]_{VR}]_{VR} [*a*]_{VS} [*r*]_V "to harden"
 b. [prefix[[NR]suffix]_{VR}]_{VR}
 [[*a*[[*pedr*]_{NR} *e*'*j*]_{VR}]_{VR} [*a*]_{VS} [*r*]_V "to stone"
 [[*en*[[*raiv*]_{NR} *e*'*c*]_{VR}]_{VR} [*e*]_{VS} [*r*]_V "to enrage"
 [prefix[[NR]]_{VR}]_{VR}
 [[*a*[[*carici*]_{NR}]_{VR}]_{VR} [*a*]_{VS} [*r*]_V "to caress"
 [[*en*[[*garra*'*f*]_{NR}]_{VR}]_{VR} [*a*]_{VS} [*r*]_V "to bottle"

Most parasynthetic forms are deadjectival (10a) or denominal verbs (10b). There are some parasynthetic adjectives too (10c):

- (10) a. $es[[clar]_{ADJR}e'c]_{VR}er$ "to clarify"
 b. $a[[cam'p]_{NR}ar]$ "to camp"
 c. $a[[laran'j]_{NR}ad]_{ADJR}o$ "orangy"

1.3. Conversion

Cases traditionally treated as back-formation and "improper" derivation fit in this category, since they both involve the recategorization of a base, without the intervention of affixes. Conversion processes are not typical morphological processes, although they have several features in common with derivational suffixation. We will mention three:

Conversion can operate on different morphological categories, namely roots (13a) and fully inflected words (13b):

- (13) a. $[ata'c]_{VR}ar \rightarrow [a'taqu]_{NR}e$ "to attack→attack"
 b. $[[[o'lh]_{VR}a]_{VS}r]_{VINFR} [o'lhar]_{NR}(es)$ "to see→look"

Conversion generates words that belong to the same morphosemantic classes as those that are formed by derivation (cf. (7), above):

- (14) causative verbs: $[lim'p]_{ADJR}ar$ "to clean"
 action nouns: $[a'taqu]_{VR}e$ "attack"
 subject nouns: $[pe'netra]_{VPI3RDSG}ns$ "intruder"

Conversion and derivation are usually in complementary distribution:⁷

- (15) $mis'tura$ vs. $*mistura'ção$ "mix"
 $*'grava$ vs. $grava'ção$ "recording"
 $melho'rar$ vs. $*melhorifi'car$ "to improve"
 $*pu'rar$ vs. $purifi'car$ "to purify"

2. Modification

Many affixation processes are of a modification kind. Morphological modifiers are adjuncts that copy grammatical features from the base they are added to and they just change its semantic value. This category includes all evaluative suffixation and all prefixation.

2.1. Evaluative suffixation

Evaluative suffixation is one of the most interesting domains in Portuguese word formation. Since it is a resource primarily used in spoken language, it is quite superficially studied in schools and no standardization is available in reference grammars. Therefore, we can see the true dynamics of these word-formation processes.

Evaluative modifiers change the base they are added to according to a range of semantic features related to different value judgments, the true content of which depends on pragmatic circumstances.⁸ Evaluative modification applies almost unrestrictedly: these suffixes can

adjoin to all kinds of bases, either adjectival (16a), nominal (16b), adverbial bases (16c), or even to interjections (16d):

- (16) a. [[[ma'gr]inh]o/a] "thin+EVAL"
 b. [[[profe'ssor]zinh]o] "teacher+EVAL"
 c. [[[ced]inh]o] "early+EVAL"
 d. [[[adeu's]inh]o] "goodbye+EVAL"

Among the existing evaluative suffixes, the linguistic variants under consideration prefer *-inho(a)* and *-zinho(a)*.⁹ In EP, *-inho* and *-zinho* are not allomorphs of a single suffix—they belong to two different series, with distinctive features: vowel-initial suffixes (henceforth V-evaluatives) are adjoined to roots; [z]-initial suffixes (henceforth Z-evaluatives) are adjoined to words. The distribution of these two sets of competing suffixes is dialect-specific and it is also prone to speaker's preference, but some grammatical constraints also apply.¹⁰

In the two varieties, the most obvious constraint is the impossibility to adjoin V-evaluative suffixes to athematic bases (i.e. roots that have identical forms for the root, the stem and the singular word ([[ca'fê]_{NR}][_{TI}][_{NS}][_{MSS}]_N "coffee"). Athematic roots only allow for Z-evaluative suffixation, displaying, in the suffix, the unmarked thematic index (=TI) that agrees with the gender of the base (17a). This is also the case for athematic roots with a stressless final vowel (17b):

- (17) a. [[[ca'fê]_{MSC}'zinh][o]_{TI}]_{MSC} *cafe'inho "coffee+EVAL"
 [[[ir'mã]_{FEM}'zinh][a]_{TI}]_{FEM} *irmã'inha "sister+EVAL"
 b. [[[táxi]_{MSC}'zinh][o]_{TI}]_{MSC} *tá'xinho "taxi+EVAL"

In some dialects of BP, *'mãe'zinha* ("mother+EVAL") may co-occur with *mã'inha* (or *'pai'zinho* and *pa'inho* "father+EVAL"), for instance. This may indicate that the constraint that holds for EP does not hold for some dialects of BP. In southern dialects of PB, these instances are felt as typical northeastern formations.

–Ø roots also show a speaker's preference for Z-evaluative suffixation (18a), which seems to indicate that there is a large proximity between –Ø roots and athematic roots. This proximity is eventually higher in BP than in EP (18b), which suggests that [l]-final roots in BP have become athematic.

- (18) a. [[['dor]'zinha] *?[[do'r]inha] "pain+EVAL"
 [[['sal]'zinho] *?[[sa'l]inho] "salt+EVAL"
 b. BP: *[[ane'l]inho]/[[a'nel]'zinho] "ring+EVAL"
 EP: [[ane'l]inho]/[[a'nel]'zinho]

In EP, we often find cases of –Ø roots in free variation, which clearly illustrate that V-evaluative and Z-evaluative suffixes attach to different bases: V-evaluatives select a root; Z-evaluatives select a word. Notice that the stressed vowel of the base gets two different phonetic realizations depending on the choice of the suffix:¹¹

- (19) *caraco'linho/cara'col'zinho* "snail+EVAL"
casa'linho/ca'sal'zinho "couple+EVAL"

In the case of the *-e* roots, chances for an equivalent distribution are even higher. It is possible to find more instances of V-evaluative and Z-evaluative suffixation that select the

same bases, in both varieties of Portuguese, as it is possible to find cases of apparently random acceptance, or non-acceptance of both, or just of one of them:

- (20) *bi'finho* *'bife'zinho* “steak+EVAL”
pei'xinho *'peixe'zinho* “fish+EVAL”
cha'vinha *'chave'zinha* “key+EVAL”

In *-a* and *-o* roots, distribution is also varied. Although it is not possible to find strict criteria, data show that the preference for V-evaluative suffixes lies in shorter highly frequent bases; the preference for Z-evaluative suffixes comes from longer and less frequent bases (which include most proparoxytonic words):

- (21) a. *bo'quinha* *ros'tinho*
'boca'zinha *'rosto'zinho*
“mouth+EVAL” “face+EVAL”
ca'rinha *de'dinho*
[?]*'cara'zinha* [?]*'dedo'zinho*
“face+EVAL” “finger+EVAL”
b. [?]*pupi'linha* *pesco'cinho*
pu'pila'zinha [?]*'pes'coço'zinho*
“pupil+EVAL” “neck+EVAL”
c. [?]*sobrance'lhinha* [?]*'crocodi'linho*
sobran'celha'zinha *croco'dilo'zinho*
“eyebrow+EVAL” “crocodile+EVAL”
d. [?]*medi'quinho* [?]*celu'linha*
'medico'zinho *'celula'zinha*
“doctor+EVAL” “cell+EVAL”

Preference for Z-evaluative suffixation¹² may be explained by the fact that Z-evaluative suffixes facilitate the recognition of the base to which they associate. Notice that Z-evaluative suffixation triggers gender and number agreement between the evaluative word and the base word, which can be clearly demonstrated when the base word has allomorphic variation for number inflection (22a). When the base is an *-a* stem masculine root, or an *-o* stem feminine root, Z-evaluative suffixation triggers agreement in gender with the base and exhibits the unmarked thematic vowel for gender: *-o* for the masculine and *-a* for the feminine. V-evaluative suffixation preserves the thematic vowel of the base (22b). Consequently, when the base is an *-a* stem root non-specified for gender, Z-evaluatives disambiguate the gender value, whereas V-evaluatives do not (22c):

- (22) a. [[[[*cará'col*]_{SG} *'zinh*o]]_{SG} [[[[*cará'coi*]_{PL} *'zinh*o]]_{PL}
[[[[*cão*]_{SG} *'zinh*o]]_{SG} [[[[*cãe*]_{PL} *'zinh*o]]_{PL}
b. [[[[*sis'tema*]_{MSC} *'zinh*o]]_{MSC} [[[[*siste'm*]_{MSC} *inh*]]_{MSC} *a*]]_{MSC} “system+EVAL”
[[[[*tribo*]_{FEM} *'zinh*a]]_{FEM} [[[[*tri'b*]_{FEM} *inh*]]_{FEM} *o*]]_{FEM} “tribe+EVAL”
c. [[[[*ar'tista*]_{MSC/FEM} *'zinh*o]]_{MSC} [[[[*artis't*]_{MSC/FEM} *inh*]]_{MSC/FEM} *a*]]_{MSC/FEM} “artist+EVAL”
[[[[*ar'tista*]_{MSC/FEM} *'zinh*a]]_{FEM}

2.2. Clipping

Clipping (or truncation) is quite productive in BP. It is a mechanism by which a word is shortened without its lexical meaning being affected, but with frequent stylistic or pragmatic nuances, which is why it is treated as a case of modification. It eliminates phonological

material at the right periphery of the base. Clippings may (23a) or may not (23b, 23c) affect morphological constituents:

- | | | | | |
|---------|----------------------|----|-----------------|---------------|
| (23) a. | <i>prole'tário</i> | >> | <i>pro'leta</i> | “proletarian” |
| | <i>comu'nista</i> | | <i>co'muna</i> | “communist” |
| b. | <i>vaga'bunda</i> | | <i>va'gaba</i> | “slut” |
| | <i>cer'veja</i> | | <i>'cerva</i> | “beer” |
| c. | <i>bijute'ria</i> | | <i>bi'ju</i> | “jewelry” |
| | <i>refrige'rante</i> | | <i>re'fri</i> | “soft drink” |

The patterns exemplified above require access to morphological and prosodic information. In (23a), we find words formed by a root base and the thematic index *-a*, a constituent unrelated to the gender of the base. In (23b), the base root is not fully present in the truncated form, but, as in (23a), the clippings are stressed on the penultimate syllable, always forming a trochee at the right edge of the shortened form. In these two groups, the affixation of the thematic index (*-a*) always takes place, but not in (23c). Here, the two first syllables of the base are kept, which form an iambic foot.

Bauer (1988: 33) questions the morphological status of such clippings because the excluded parts are not clearly morphological. According to Fandrych (2008: 116), clipping is unquestionably a process of word formation: the shortening “changes records or styles compared to their complete counterparts.” This is indeed what is observed in the examples (23a–b), the most common cases. This pattern of clipping can also affect compounds:¹³

- | | | |
|------|-------------------------|--------------------------------|
| (24) | <i>'São Gon'ç(alo)a</i> | “a district of Rio de Janeiro” |
| | <i>'grã-'f(ino)a</i> | “snobbish” |
| | <i>'free-'l(ancer)a</i> | “freelancer” |

2.3. Prefixation

Like evaluative suffixes, modifiers that are left-adjoined to a head do not interfere with the grammatical properties of the words in which they occur:

- | | | | |
|------|-----------------|--------------------|----------------------|
| (26) | <i>'apto_</i> | <i>[i'n]apto</i> | “fit→unfit” |
| | <i>fa'zer_</i> | <i>[des]fa'zer</i> | “do→undo” |
| | <i>ma'rido_</i> | <i>[ex]ma'rido</i> | “husband→ex-husband” |

The range of semantic values expressed by prefixes is wider than for evaluative suffixes. Prefixation can also be evaluative (*[super]interes'sante* “super-interesting,” *[micro]computa'dor* “microcomputer”), but there are prefixes of negation (*[in]e'quívoço* “unequivocal”), opposition (*[des]mon'tar* “dismount”), repetition (*[re]encon'trar* “meet again”) and spatial (*[sub]solo* “subsoil”) or temporal location (*[pós]opera'tório* “post-surgery”).

Regardless of the semantic diversity found amongst prefixes, it is important to remark that the set of units usually called prefixes may have very different properties. The heterogeneity of these units can be analyzed according to a number of criteria showing that some behave like typical affixes, while others look like independent roots.¹⁴

The first criterion is related to the category of the base to which they may attach: some prefixes attach exclusively to roots or stems (27a), while others, especially those that can be coordinated with other prefixes, attach to words or even to phrases (27b).

- (27) a. [[[in][apt]]i'dão] "inaptitude"
 [[des][arm]a]'mento] "disarmament"
 [[re][aprecia]]'çãõ] "re-appreciation"
 b. [['pré][cam'panha]] "pre-campaign"
 [['pré][cam'panha eleito'ral]] "pre-electoral campaign"
 [['pré e 'pós][cam'panha (eleito'ral)]] "pre- and post-(electoral) campaign"

Prefixes that attach to roots or stems (typically *in-*, *des-* and *re-*) are unstressed units; those that adjoin to words or syntactic expressions (such as *pós-*, *pré-*, *ex-* and *sub-*) are independent prosodic words (cf. Schwindt 2000). The latter type is formed by paroxytones with two syllables ('*contra-*', *an'ti-*, '*mega-*', and '*super-*') or stressed monosyllables ('*pró-*', '*ex-*', and '*pré-*').

One issue that must be raised in relation to this distinction concerns prefixes that have stressed and unstressed variants, like '*pré-/pre-*' or '*pós-/pos-*'. For instance, in (28a) the prefix is unstressed ([pri] in EP, [pre] in BP); in (28b) the prefix is stressed ([pré] in both varieties):

- (28) a. [[pre]['texto]] "excuse"
 [[pre][s'sentir]] "to sense"
 b. [['pré][pro'jeto]] "pre-project"
 [['pré][da'tar]] "to predate"

There is an obvious difference of formal and semantic transparency between words in (28a) and those in (28b): the former are opaque and can be seen as genuine cases of lexicalization—*pre'texto*, for example, means "excuse," which is not related to *texto* ("text") nor to a temporal location value of the prefix. Thus, stressed prefixes become unstressed when the words get lexicalized: either for semantic reasons (*pres'sentir* "to sense") or for structural reasons, when the prefix is adjoined to a neoclassical bound root (*prema'turo* "premature").

A second criterion that is relevant to isolate prefixes concerns their (in)existence as autonomous words. Many of these forms are derived from Greek or Latin prepositions and adverbs, which have undergone a process of grammaticalization already in the old languages. In some cases, these prefixes only occur in lexicalized words ([a]ssu'*mir* "to assume," [con]su'*mir* "to consume," [pre]su'*mir* "to presume," [re]su'*mir*, "to summarize") and therefore their historical origin is irrelevant for their synchronic classification.

However, Portuguese follows the model of the classical languages by using prepositions and adverbs to build modified words:

- (29) [['ante]['câmara]] "antechamber"
 [[sem][a'brigo]]/[[sem]['teto]] "homeless"
 [['não][agres'são]] "non-aggression"

Also available for this type of modification are neoclassical forms, which may have served as prefixes in old languages, and which are again available in contemporary Portuguese (as in many other modern languages):

- (30) a. [['hiper][a'tivo]] "hyperactive"
 [['sobre][do'tado]] "overly gifted"
 b. [['hemi][atro'fia]] "semi-atrophy"
 [['meio][ir'mão]] "half-brother"

The third criterion distinguishes forms that can only occur as left adjuncts from those that can themselves be the head of a complex word. The first class are prefixes (31a), the second one are roots (31b):

- (31) a. [[*'mega*][*manifesta'çã*o]] “huge demonstration”
 b. [[[*pat*][[*o*][*lo'g*]]*ia*] “pathology”
 [[[*cardi*][[*o*][*pa't*]]*ia*] “heart disease”

Thus, both unstressed forms adjoined to roots or stems and stressed forms adjoined to words or phrases have a similar behavior, which raises the question of whether it is possible to find independent grammatical evidence for their different categorization.

2.4. Productive phonology in affixation

Root- and stem-based derived and modified words behave like simple words with respect to stress assignment: they always form a single stress domain. Consequently, in these cases, there is an isomorphism between morphological and prosodic words. Looking at the effects of unstressed vowel reduction processes in EP and BP helps to sustain this claim:

In EP, all low and mid vowels, before and after stressed syllables, undergo reduction and centralization:

- (32) *'v[ɛ]la* *v[i]'leiro* “sail/sailboat”
'b[e]rço *b[i]r'çário* “cradle/nursery”
'b[a]rco *b[ɐ]r'queiro* “boat/boatman”
's[ɔ]l *s[u]'lar* “sun/solar”

In BP, only mid vowels are concerned: lower mid vowels of the base alternate with their corresponding upper mid vowels when they occur in a pre-stress position:

- (33) *'p[ɔ]rta* *p[o]r'teiro/p[o]rta'ria* “door/doorman/hallway”
'v[ɛ]la *v[e]'leiro/v[e]le'jar* “sail/sailboat/to sail”

Root based evaluative words behave differently. In EP, mid vowels and low central vowels are always reduced (34a); palatal and velar low vowels are preferably not reduced, although reduction may occur (34b):

- (34) a. *'d[e]do* **d[e]'dinho/* *d[i]'dinho* “finger(+EVAL)”
'b[o]lo **b[o]'linho/* *b[u]'linho* “cake(+EVAL)”
'c[a]sa **c[a]'sinha* *c[ɐ]'sinha* “house(+EVAL)”
 b. *'f[ɛ]sta* *f[ɛ]s'tinha* *f[i]s'tinha* “party(+EVAL)”
'b[ɔ]la *b[ɔ]'linha* *?'b[u]'linha* “ball(+EVAL)”

In evaluative words all vowels keep their underlying quality in BP. Compare the mid vowel quality of the examples below with their correspondents in derived words:

- (35) *'p[ɔ]rta* *p[ɔ]r'tinha* “door(+EVAL)”
 p[o]r'teiro/p[o]rta'ria “doorman/hallway”
'v[ɛ]la *v[ɛ]'linha* “sail(+EVAL)”
 v[e]'leiro/v[e]le'jar “sailboat/to sail”

Word-based suffixed words form two stress domains, which means that the isomorphism between morphological and prosodic words is broken: one morphological word projects into

we claim that this behavior follows from a selectional property of the prefixes, which may also attach to roots or to words: when they attach to roots they expand the prosodic word of the base ($d[i]s-$, $r[i]$); when they attach to words they project a new prosodic word ($n[\varepsilon]o-$, $pr[\varepsilon]-$, $p[\varepsilon]s-$).

Notice that, in BP, *re-* and *pre-*, unlike *des-*, do not undergo vowel harmony. It leads to the raising of unstressed mid vowels, under the influence of a following stressed high vowel, similar to what happens with unprefixated forms such as *pe'pino* ("cucumber"), *pre'guiça* ("laziness") or *sen'tir* ("to feel")—they may be pronounced with [i] in the initial syllable (see Bisol and Veloso, Chapter 5 in this volume, for details). For instance, in BP, phonetic forms such as $*r[i]vi'sita$ ("revisit"), $*r[i]tinto$ ("re-dye") and $*r[i]visto$ ("revised") do not occur, nor do $*pr[i]ver$ ("to preview") e $*pr[i]s'into$ ("I sense"), which suggests the existence of some morphological conditioning for the application of this phonological rule. Moreover, the raising of the stressed vowel does not occur in hiatus ($*r[i]abaste'cer$ "to refill"; $*r[i]u'sar$ "to reuse"), and the vowels in these prefixes are less likely to deletion when the following vowel is identical ($r[\varepsilon]-[e]xpli'car$ "re-explain," $r[\varepsilon]-[e]labo'rar$ "re-elaborate").¹⁶ Despite being unstressed, this prefix has a high degree of morphological integrity, since it resists several processes that affect the unstressed vocalism.

Finally, we mention the neutralization of mid vowels in stressed syllables. Suffixes like *-ico* and *-il* form dactylic (41a) and spondaic feet (41b), respectively. Stressed mid vowels in these derivatives always display lower mid qualities ($[\varepsilon]$, $[\varepsilon]$), which is also the pattern for non-derived words with the same prosodic structure (41c). The productivity of the process is shown by the fact that loans adapt to these models ($'W[\varepsilon]shington$; $'sh[\varepsilon]pping$). These facts prove that lowering is a phonological rule in Portuguese (cf. Wetzels (1992)).

- (41) a. *esque'I[ε]to* *esque'I[ε]tico* *ca'I[o]r* *ca'I[ε]rico*
 "skeleton" "skeletal" "heat" "caloric"
'núm[ε]ro *nu'm[ε]rico* BP: *me'táf[o]ra* *meta'f[ε]rico*
'núm[i]ro *nu'm[ε]rico* EP: *me'táf[u]ra* *meta'f[ε]rico*
 "number" "numerical" "metaphor" "metaphorical"
- b. BP: *proj[ε]'tar* *pro'j[ε]til* *'d[o]ce* *'d[ε]cil*
 EP: *proj[ε]'tar* *pro'j[ε]til* "sweet" "docile"
 "to project" "projectile"
- c. *'p[ε]tala* *'f[ε]sforo* *'r[ε]plica* *'c[ε]cegas*
 "petal" "match" "replica" "tickles"

2.5. Affixal allomorphy

Some derivational suffixes have one or more allomorphs. In some cases, like *-al* ~ *-ar*, for instance, this allomorphy was inherited from Latin: *-ar* occurs due to a dissimilation when the nominal base contains /l/:¹⁷

- (42) a. $[[aciden't]al]$ "accidental"
 $[[horizon't]al]$ "horizontal"
 b. $[[celu'I]ar]$ "cellular"
 $[[molecu'I]ar]$ "molecular"
 c. $[[elemen't]ar]$ "basic"
 $[[nucl'e]ar]$ "core"

Another case of allomorphy (which does not have a Latin origin) concerns the suffix *-ez* ~ *-eza*,¹⁸ that forms deadjectival quality nouns:

- (43) *'velh(o)* *ve'lhice* ~ *ve'lhez* "old/old age"

Contemporary formation of quality nouns prefers another suffix (i.e. *-idade*), which means that *-ez* ~ *-eza* is not used to form new words. Nevertheless, available data indicate that their distribution has a prosodic basis, which is related to their thematic status: *-ez* forms \emptyset stem nouns, such as *timi'dez* "shyness," *-eza* forms *-a* stem nouns, like *ma'greza* "slimness." The allomorph distribution is sensitive to the number of syllables in the base:¹⁹ *-eza* selects shorter bases (typically monosyllable roots) such as *fri'eza* ("coldness"), and *-ez* selects longer bases (roots with two or more syllables), like *aci'dez* ("acidity").

Prefixes can also have allomorphic variation. The prefix *in-*, for instance, has three allomorphs: [i.n], [i] e [ĩ]:

- | | | |
|--------------------------|---------------------------|----------------------------------|
| (44) [i]le'gal "illegal" | [ĩ]pos'sível "impossible" | [i.n]apto "unfit" |
| [i]mo'ral "immoral" | [ĩ]'certo "uncertain" | [i.n]experi'ente "inexperienced" |
| [i]rre'al "unreal" | [ĩ]'justo "unfair" | [i.n]útil "useless" |

These alternations are the same in EP and BP: [i] occurs before sonorant consonants, [ĩ] is chosen before a base-initial non-sonorant, whereas the sequence [in] is found before a vowel-initial base.

Some derivational suffixes trigger the application of morphophonological rules that affect the phonetic shape of the output, such as the lenition of velars (45a) or haplology²⁰ (45b):

- | | | | |
|---------|----------------------|---------------------------------|--------------------------|
| (45) a. | <i>fi'lólo</i> [g]o | <i>filolo</i> '[ʒ]ia | "philologist/philology" |
| | <i>his'tóri</i> [k]o | <i>histori</i> '[s]ismo | "historic/historicism" |
| b. | <i>cari'dade</i> | <i>cari(da)</i> ' <i>do</i> so | "charity/charitable" |
| | ' <i>mínima</i> | <i>se(mi)</i> '(mi) <i>nima</i> | "half note/quarter note" |

In the first case (cf. Lee 1995), velar plosives become fricatives before suffixes initiated by the vowel /i/, like *-ia*, *-ista* and *-ismo*. In morphological haplology, two identical or phonetically similar syllables are reduced to one, usually the right one (cf. Gonçalves 2011).

3. Compounding

Affixation requires the presence of an affix, compounding combines roots or words: the combination of roots yields a morphological structure (a morphological compound); the combination of words yields lexical units that have a hybrid morphosyntactic structure (morphosyntactic compounds) or they are lexicalized phrases (syntactic compounds).

We have just seen that the difference between compounding and affixation suggests that there is a clear-cut distinction between affixes and roots (or even words), which, in fact, does not exist. We will nevertheless present a characterization of roots that helps to set them apart from affixes.

3.1. Roots

In Portuguese, some roots occur in simple and in complex words, as [metr] in ['metr]o "meter," [[['métr]ic]o] "metric" and [[[me'tr][ónom]]]o "metronome"; other only occur in complex words, as ([fratr] in [[[fratr][i]['cid]]]a "fratricide"). Derivational suffixes (like *-ção*) and modifier affixes (like *des-*) can only occur in complex words (*liga'ção* "connection"; *desli'gar* "to disconnect"). Thus, it is easy to distinguish roots (that may occur in both simple and complex words) from affixes (that cannot occur as simple words); it is harder to set apart roots that occur only in complex words, like affixes.

Roots that are present in simple words are roots by definition. Simple words make their inherent properties (such as word-class and subcategories (45a)) visible. These roots can also occur in words formed by derivation or modification (45b):

- (45) a. $[[[b'ol]_{NR-A, FEM} a]_{NS^4 N}]$ "ball"
 $[[[b'ol]_{NR-O, MSC} o]_{NS^4 N}]$ "cake"
 $[[[ge'r]_{C1 VR} a]_{VS} r]_V$ "to generate"
 $[[[ge'r]_{C3 VR} i]_{VS} r]_V$ "to manage"
 b. $[[[[b'o'l]_{NR} ad]_{NR} a]_{NS^4 N}]$ "hit with a ball"
 $[[[[b'o'l]_{NR} inh]_{NR} o]_{NS^4 N}]$ "small cake"
 $[[[[ger]_{VR} a' dor]_{NR}]_{NS^4 N}]$ "manager"
 $[[[[ge'r]_{VR} ent]_{NR} e]_{NS^4 N}]$ "manager"

Roots that cannot occur in simple words are generally loans from classical languages that are particularly productive to form technical terms. These roots occur mainly in morphological compounds (46a), but they can also be selected by neoclassical suffixes (46b). Usually, they have an imprecise meaning, and they are underspecified for word-class and thematic membership. They depend on other constituents to become a member of a word-class (46c):

- (46) a. $[[[top][o][lo'g]]ia]$ "topology"
 b. $[[[t'op][ic]]o]$ "topic"
 c. $[[[bi][o][lo'g]]ia]_N$ "biology"
 $[[[bi][o][l'og]]ico]_{ADJ}$ "biologic"

Many of these roots can take any of the available positions (47a), but there are roots that can only be in the initial position (47b) and others that can only occur in final position (47c):

- (47) a. $[antro'p]ólogo$ "anthropologist"
 $fil[an'trop]o$ "philanthropist"
 b. $[hom]o'nímia$ "homonymy"
 c. $herbi[cid]a$ "herbicide"

In the previous section, we established that roots that occur only as initial constituents, like *hom-* (a loan from Greek, meaning "equal") and prefixes are better described, indistinctly, as modifiers. The classification of units like *antrop* (a loan from Greek, meaning "man") as roots, rather than affixes, derives from their availability in both initial and final position. Finally, roots that can only occur in the final position of morphological compounds, such as *-cid* (a loan from Latin, meaning "kill") are different from derivational suffixes, because they do not define the word-class of their output,²¹ and they are different from evaluative suffixes because they are heads.

3.2. Morphological compounds

Morphological compounds may have a modification or a coordination structure. Modification structures are the result of left adjunction of a root (the modifier) to another root (the head):

- (48) $[[pat]_{MODIFIER} o[lo'g]_{HEAD}]ia$ "pathology"
 $[[en'cefal]_{MODIFIER} o[pa't]_{HEAD}]ia$ "encephalopathy"

In coordination structures, both roots are heads:

- (49) [¹*crani*][*o*][*ence*'*fál*][*ico*] “cranioencephalic”

Usually, the boundary between roots is marked by the binding vowel *-o-*, unless it precedes a member of a (lexically determined) small set of Latinate roots, and only in modification structures. In this case, the binding vowel is *-i-*:

- (50) *hom*[*i*]'*cida* “homicide”
frut[*i*]*cul*'*tura* “fruit production”
ver'*m*[*i*]*fugo* “vermifuge”
ampl[*i*]'*forme* “ampliform”
car'*n*[*i*]*voro* “carnivore”

The binding vowel may be absent in modification structures (not in coordination structures). This absence occurs when the right-hand root begins with a vowel:²²

- (51) *dem*[*]*'*ago*'*gia* “demagogy”
'*sul*[*]*'*ameri*'*cano* “South American”

From a morphological point of view, these compounds are tripartite structures (root-binding vowel-root). Since the choice of the binding vowel is sensitive to the kind of structure (modification vs. coordination) and to a lexical feature of the right-hand root, the binding vowel is a morphological specifier of the right-hand root. From a phonological/prosodic point of view, morphological compounds project two prosodic words and the binding vowel is the final vowel of the first prosodic word.

The phonetic realization of the binding vowel is quite interesting in itself and it is also quite revealing. Binding vowels are usually in an unstressed position—that changes whenever the right-hand root’s only vowel cannot be stressed and no derivational suffix is present: In BP the stressed mid vowel is lower mid in these words (dactylic lowering):

- (52) *au*'*t*[*ó*]*grafo* “autograph”
bibli'[*ó*]*filo* “bibliophile”
psi'*c*[*ó*]*logo* “psychologist”
ver'*m*[*í*]*fugo* “vermifuge”

These cases are irrelevant for the analysis of the quality of the binding vowel. On the contrary, the quality of the binding vowel *-o-* when it is in a non-stressed position is worth a note. In EP, this vowel resists to unstressed vowel reduction, surfacing as [ɔ] (53a), unless the word is lexicalized, which means that it becomes a single prosodic word (53b). In BP, the binding vowel tends to be surfacing as [u] (57a). In lexicalized instances, the binding vowel also surfaces as a different vowel ([o]): (looks as if the underlying value is lower mid in EP and upper mid in BP).

- | | | | |
|------|---|---|---------------------|
| (53) | EP: | BP: | |
| a. | <i>p</i> ' <i>sic</i> [<i>ɔ</i>] <i>lin</i> ' <i>guística</i> | <i>p</i> ' <i>sic</i> [<i>u</i>] <i>lin</i> ' <i>guística</i> | “psycholinguistics” |
| | ' <i>aut</i> [<i>ɔ</i>] <i>susten</i> ' <i>tável</i> | ' <i>aut</i> [<i>u</i>] <i>susten</i> ' <i>tável</i> | “self-sustained” |
| | ' <i>micr</i> [<i>ɔ</i>]' <i>clima</i> | ' <i>micr</i> [<i>u</i>]' <i>clima</i> | “microclimate” |
| b. | <i>fil</i> [<i>u</i>] <i>so</i> ' <i>fia</i> | <i>fil</i> [<i>o</i>] <i>so</i> ' <i>fia</i> | “philosophy” |
| | <i>aut</i> [<i>u</i>] <i>gra</i> ' <i>far</i> | <i>aut</i> [<i>o</i>] <i>gra</i> ' <i>far</i> | “to autograph” |
| | <i>micr</i> [<i>u</i>] <i>s</i> ' <i>cópio</i> | <i>micr</i> [<i>o</i>] <i>s</i> ' <i>cópio</i> | “microscope” |

Some morphological compounds use clips from other morphological compounds. As neoclassical loans, they convey their original meaning (54a), but as clips they bring the overall meaning of the compounds from where they originated (54b):

- (54) a. 'fotos'síntese "photosynthesis"
 fotogra'fia "photography"
 'autorre'trato "self portrait"
 auto'móvel "car"
 biogra'fia "biography"
 biolo'gia "biology"
 econo'mia "economy"
 ecolo'gia "ecology"
- b. foto('grafia)+jorna'lista "photo+journalist"
 'fotojorna'lista "photographic journalist"
 auto('móvel)+es'trada "car+road"
 'autoes'trada "freeway"
 bio(lo'gia)+degra'dável "biology+degradable"
 'biodegra'dável "biodegradable"
 eco(lo'gia)+tu'rismo "ecology+tourism"
 'ecotu'rismo "ecological tourism"

Clips such as *agro-*, *bio-*, *eco-*, *eletro-*, or *foto-* become new roots by merging the binding vowel with the neoclassical root. Consequently, they have a different behavior: for instance, when they precede a vowel-initial root, the final *-o* is not deleted:

- (55) foto(gra'fia)+aven'tura "photo+adventure"
 'fotoaven'tura "photographic adventure"
 eco(lo'gia)+al'deia "ecology+village"
 'ecoal'deia "ecological village"

Clips often become words. That is the case of *foto* ("photo") and *micro* ("microphone"). BP has many more examples:

- (56) Meus irmãos são 'héteros. (=hetero[ssexu'ai]s)
 "My brothers are heterosexuals"

3.3. Morphosyntactic compounds

Morphosyntactic compounds result from the right adjunction of a noun to a noun (57a), the coordination of nouns (57b) or, less frequently, of adjectives (57c) or even of verbs (57d), and the reanalysis of a verb phrase (57e):

- (57) a. 'mãe-co'ruja "doting mother"
 b. lei'tor-grava'dor "player-recorder"
 c. 'surdo-'mudo "deaf-mute"
 d. 'leva-e-'traz "gossiper"
 e. 'quebra-'nozes "nutcracker"

Modification structures like those in (57a) and (58) are head-initial: the head noun determines the gender and the number of the compound. The modifier noun remains invariable:

- (58) a. [[ca'fé(s)]_{N_MSC_SG/PL} con'certo]_{N_MSC_SG} "coffee concert = cabaret(s)"
 b. [[es'cola(s)]_{N_FEM_SG/PL} mo'delo]_{N_FEM_SG} "school model = model school(s)"

These compounds are interpreted as modification structures: the compound is a hyponym of its head: *ca'fé-con'certo* is a “type of coffee house”; *es'cola mo'delo* is a “type of school.”

In coordination structures like (59b) and (59c), both constituents are heads. Therefore, they both inflect in number (59a). For animated nouns (59b) or adjectives (59c), gender agreement is also required. In the case of coordination of inanimate nouns with discordant values of gender, the gender is masculine (59d), which is the unmarked value:

- (59) a. [*lei'tor(es)*]_{N_MSC_SG} [*grava'dor(es)*]_{N_MSC_SG} “player-recorder(s)”
 b. [*nada'dor/a(s)*]_{N_MSC/FEM_SG/PL} [*salva'dor/a(s)*]_{N_MSC/FEM_SG/PL} “lifeguard(s)”
 c. [*doce(s)*]_{ADJ_SG} [*a'margo/a(s)*]_{ADJ_MSC_SG} “bittersweet”
 d. [*bar(es)*]_{N_MSC_SG} [*disco'teca(s)*]_{N_FEM_SG} “disco-bar(s)”

The meaning of coordinated compounds is not always easy to establish, since it may be additive (60a), sequential (60b), or reciprocal (60c):

- (60) a. *lei'tor-grava'dor* “reader-recorder = device that plays and records”
 b. *ou'tono-in'verno* “autumn-winter”
 c. *a'luno-profes'sor* “student-teacher = relationship between student and teacher”

Note that the line between modification structure and coordination structure may be difficult to draw: in some cases they may be interpreted as reciprocal modification (61a). This difficulty is noticeable in the hesitation of speakers regarding number inflection (61b):

- (61) a. *so'fá-'cama* “sofa-bed” = sofa that serves as a bed or a sofa that serves as a bed and a bed that serves as a sofa
 b. *so'fás-'camas* vs. *so'fás-'cama*

Another type of morphosyntactic compounding is based on structures very similar to VPs. They are formed by the third-person singular indicative present form of a (typically) transitive verb; and by a (generally) plural noun, which is the head of the direct object of that verb. Usually, this compounding process generates a masculine subject-noun:

- (62) '*guarda-'costas* ‘guards-back = bodyguard’
 '*quebra-ca'beça(s)* ‘breaks-heads = puzzle’

A final type of morphosyntactic compounding combines two verbs to form, again, a masculine noun. Two subtypes must be distinguished. The first corresponds to the coordination of two different verb forms (V_iV_i). Usually, these compounds occur only in the singular form and their meaning is quite transparent:

- (63) BP: '*bate-en'tope* ‘hits-clogs = hit-clog’
 EP: '*sobe-e-'desce* ‘goes up-and-comes down = see saw’
 BP/EP: '*leva-e-'traz* ‘takes-and-brings = intriguer’

The second subtype, much more common in BP than in EP, involves reduplication of the verb to form a V_iV_i compound. These forms can convey two meanings: an action (64a) or an object (64b). In some cases, both meanings can be observed in the same word (64c):

- (64) a. '*corre-'corre* ‘run-run’
 b. '*pisca-'pisca* ‘blinks-blinks = blinker’
 c. '*pula-'pula* ‘jumps-jumps =act of jumping repeatedly/a toy in the playground’

The bases of V_1V_1 compounds are generally disyllabic. There are also a few cases like *a'garra-a'garra* (“grabs-grabs”), with three syllables that always start with an onsetless syllable. Since the reduplication of the verb base is governed by prosodic conditions, the final syllables are always open, except when the verb is monosyllabic (*'sai-'sai* “leaves-leaves” = “one goes out”). Finally, the main morphological characteristic of this type of formation is the selection of the third-person singular indicative present: we assume that this is the unmarked form of the verb paradigm, which allows the reinterpretation of the verb as a noun.

3.4. Syntactic compounds

Syntactic compounding is not a morphological word formation process—it is a process of lexicalization of phrases. Reference grammars usually list a number of different cases, such as the following:

- | | | | |
|------------|--|-------------------------|---------------------------|
| (65) N-P-N | [[N] _P [[N] _{NP}] _{PP}] _{NP} | <i>ca'minhode'ferro</i> | “road-of-iron = railroad” |
| N-ADJ | [[N][ADJ] _{ADJP}] _{NP} | <i>'cofre'forte</i> | “safe-strong = safe” |
| ADJ-N | [[ADJ] _{ADJP} [N] _{NP}] | <i>'altatempo'rada</i> | “high season” |

These word sequences display typical syntactic properties, regarding number inflection (66a) and gender contrasts, when available (66b):

- | | | |
|---------|----------------------------------|--------------------------|
| (66) a. | <i>ca'minho(s)de'ferro</i> | “railroad(s)” |
| | <i>'cofre(s)'forte(s)</i> | “safe(s)” |
| b. | <i>arqui'teto/adeinteri'ores</i> | “interior(m/f) designer” |
| | <i>pri'meiro/ami'nistro/a</i> | “prime minister (m/f)” |

What motivates the treatment of these sequences as lexical units is their semantics, which is not compositional. Another property that distinguishes these lexicalized phrases from genuine syntactic phrases is the fact that the extraction of a single constituent is ungrammatical:

- | | | |
|--------|---|---------------------------------------|
| (67) * | <i>caminhos de [ferro velho]</i> | “[old iron] roads” |
| | <i>*esseé um [cofre quê]?</i> | “this is a [safe what]?” |
| | <i>*[primeiro e único] ministro</i> | “prime and only minister” |
| | <i>*dos dois[caminhos_i, prefiro o [_i] [de ferro]_i]</i> | “of both ways, I prefer the iron one” |

3.5. Other types of compounding?

A non-concatenative morphological process often associated with compounding is blending. Although there are two words that serve as input for a third form (as in compounding), blends are different, because they are produced by the intersection of bases instead of concatenation, as in *cren'tino* (*'crente+cre'tino*, “religious+fool” = “false religious”) and *lixera'tura* (*'lixo+litera'tura*, “garbage+literature” = “shoddy literature”). The deleted material is not predictable.

From a phonological perspective, blends are single prosodic words. The output form preserves the largest possible number of identical segments of the input forms, as in *aperta'mento* “small apartment” (*a'perto+aparta'tamento* “clench+apartment”). As a result, the transition of the first of the input forms to the second coincides with an identical segment or syllable (*'saco+pico'lé* “bag+popsicle” = *saco'lé* “popsicle in a bag”; *'pai+'mãe* “father+mother” = *'pãe* “caring father”).

3.6. Emergence of new morphological constituents

The emergence of a new productive word-formation process may happen when speakers start using a loan to make new words, or because speakers reinterpret an existing morpheme or part of a morpheme with a new meaning. In the first case, the use of formatives such as *cyber-*, *wiki-*, and *e-*, which, combined with native bases, form words like '*cyber-a'vó*' ("cyber-grandmother"), '*wiki-'aves*' ("wiki-birds") and *e-profes'sor*' ("e-professor"). In the second case, the phenomenon can be seen as a kind of folk etymology: words without any internal structure may be reinterpreted as compounds or affixing forms that consist of two parts, like *ma'drasta*' ("stepmother") and *patro'cínio*' ("sponsorship"). They are intentionally misanalyzed as *má*' ("bad") plus *drasta* to form a new meaning: "a bad stepmother"²³ and '*pa(i)*' ("father") plus *trocínio*' ("sponsored by the father"). This strategy gives rise to forms such as:

- (68) a. '*sogra'drasta* "stepmother-in-law"
 '*ir'mã'drasta* "step-sister"
 b. '*tiotro'cínio* "sponsored by an uncle"
 '*mãetro'cínio* "sponsored by the mother"

These particles are usually called splinters, which are elements that occur at the edge of the word, the same way affixes do, but, because of their meanings, they correspond to roots. Therefore, splinters form a separate class, situated somewhere between roots and affixes. Thus, clipping and blending play an important role in the morphology of Portuguese, as they can produce splinters, not being, therefore, interpreted as exclusively non-morphemically.

4. Conclusion

The study of Portuguese morphology largely benefits from the fact that (at least) two subsystems can be easily compared: contrasts between EP and BP often offer the possibility to consolidate analyses independently outlined or, inversely, show how closely related languages may be different and ask for different analyses for certain subsystems of the grammar.

The vowel system in EP, which facilitates the identification of unstressed reduced vowels, is particularly relevant to establish that suffixes may attach to roots, stems or words, and it is also required to understand the structure of morphological compounds.

The description of processes that involve non-concatenative morphology or structural mutations clearly benefit from the livelihood they have in BP. In this variety, process like reduplication, blending and clipping are most commonly employed, which enables us to say that this is one of the main aspects that differentiate the two varieties described here.

NOTES

- 1 We will discuss the formation of nouns, adjectives and verbs and *-mente* adverbs. In all examples, the stress mark (') precedes the stressed syllable. Moreover, these syllables are written in bold.
- 2 Thematic constituents (thematic vowels for verbs and thematic indexes for all other classes) are morphological specifiers.
- 3 Morphosyntactic specifiers (= MSS) are inflectional suffixes.

- 4 Alongside masculine (= MSC) and feminine (= FEM), we will consider a third value (non-specified) that is assigned to bases that will be syntactically specified:

a'*tletas*_{SUB} 'athletes'
*os*_{MSC}'*a*'*tletas*_{MSC}
*as*_{FEM}'*a*'*tletas*_{FEM}

- 5 Consider, for instance, the above-mentioned inflection of '*pele*' ('skin') and '*pa'pel*' ('paper'). The relevance of thematic classes for inflection is also obvious in the phonetic outputs of words ending in the diphthong [aw]. In BP, words ending in <l>, as '*ca'nal*' ('channel'), are pronounced the same way as words ending in <u>, as '*de'grau*' ('step'). In the plural, however, they differ considerably, since the plural of the latter, which is a projection of an athematic root, is obtained by the adjunction of the suffix *-s* ('*de'graus*', 'steps'), while the plural of the former is obtained by semi-vocalization of the final consonant of the *-Ø* root, /l/, to receive the plural suffix ('*ca'nais*' 'channels').
- 6 Traditional accounts do not usually mention infixation.
- 7 There is a limited number of pairs of words formed by conversion and derivation from a same base, but generally they are not semantically equivalent ('*perda*' 'loss' and '*perdi'ção*' 'perdition'; '*celebrar*' 'to celebrate' and '*celebri'zar*' 'to make famous').
- 8 The examples presented in this section are usually called diminutive suffixes. Their semantic role may be related to size, but it may also convey other meanings: '*ca'sinha*' (the 'diminutive' from '*casa*' 'house') may refer a 'small house', a 'lovely house', a 'cherished house', an 'old house' an 'ugly house', or other, depending on pragmatics. It may even be used as a rhetoric resource. This is why the tag 'evaluative' (including augmentative and superlative as well) seems more appropriate—it includes morphological devices that allow the speaker to convey an opinion about a lexical unit, from its inside.
- 9 Portuguese southern dialects prefer *-it(o/a)* ('*li'vrito*' 'book+EVAL'), and *-zit(o/a)* ('*cão'zito*' 'dog+EVAL').
- 10 Other analyses are available, (see Bisol 2010, for instance, which considers *-inho* and *-zinho* allomorphs of the same morpheme). We follow the proposal of Villalva (1994, 2008, 2009).
- 11 In BP, *-Ø* roots that may combine with *-inho* are those ending in <r>, although few examples are found: '*colhe'rinha*'/'*colher'zinha*' 'spoon+EVAL', '*devaga'rinho*'/'*devagar'zinho*' 'slowly'. Other instances are lexicalized words ('*cola'rinho*' 'collar'+EVAL = foam of beer').
- 12 This preference has been demonstrated by the results of usage surveys (Villalva 2009).
- 13 In EP, there are cases of truncation, like '*a'narca*' ('anarchist'), which have a clear pejorative or derogatory value, but their prevalence in EP is smaller than in BP. Evidently, this is because in EP word-shortening is generally obtained by reducing unstressed vowels.
- 14 Most authors (e.g. Basílio 1987, Sandmann 1989) consider prefixation a derivational process. Some others (e.g. Mattoso Câmara 1971, Macambira 1978) argue that there are no substantial differences between prefixed and compound words; other still (e.g. Villalva 1994, Gonçalves 2012) argue that prefixation is midway between derivation and compounding. Our claim here is that prefixation is neither derivation nor compounding—it is a process of morphological modification that can make use of prefixes or roots.
- 15 In words such as '*nj]lolo'gismo*' ('neologism'), '*pr[i]ssenti'mento*' ('feeling'), or '*p[u]s'por*' ('to post pose'), the prefix vowel is reduced as a result of lexicalization.
- 16 In EP, these hiatus are avoided by glide formation ('*r[j]labaste'cer*' 'to refill', '*r[j]u'sar*' 'to re-use'), and the vowels in these prefixes are less likely to deletion when the following vowel is identical ('*r[j]-[v]xpli'car*' 'to re-explain', '*r[j]-[e]labo'rar*' 'to re-elaborate').
- 17 Words such as '*colegi'al*' ('collegial'), '*coloni'al*' ('colonial') and '*coloqui'al*' ('colloquial') do not respect the *-ar* ~ *-al* allomorphy. This is probably due to the fact that they are Latinate words recently introduced in the Portuguese lexicon.
- 18 This suffix is etymologically related to the suffixes *-ice* ('*doi'dice*' 'insanity'), and *-ície* ('*imun'ície*' 'filth'), but these behave as different suffixes, not as allomorphs, in Portuguese.
- 19 This suffix is no longer productive. Consequently, all the derivatives tend to be lexicalized. Therefore, we find the same base with both allomorphs ('*du'rez*', '*du'reza*' 'hardiness'), although usually only one of them is currently used. Some counter-examples, such as '*ru'dez*' ('rudeness') or '*aspe'reza*' ('roughness'), can also be found.

- 20 Haplogy cases are quite rare and most of them are quite old.
- 21 The word-class of *herbi'cida* (adjective/noun) is a property of the structure.
- 22 The binding vowel is not deleted when the first root ends in a vowel (*bi+log* 'life+knowledge' = *bi[o]logia* 'biology'; *ge+graf* 'earth+write' = *ge[ô]grafo* 'geographer'). There is a considerable number of morphological compounds that do not fit in the above-described pattern. Either because of a preference for the most usual binding vowel (*parc[ô]metro* 'parkmeter') or for different reasons, which may be etymological, in some cases, contrastive, in some other, since most of these words are loans that can be found in many European languages.
- 23 In BP, a loving stepmother is named *mãe'drastra* ('*mãe+ma'drastra* 'mother+stepmother'), another lexical blend.

REFERENCES

- Basílio, M. (1987). *Teoria Lexical*. São Paulo: Ática.
- Bauer, L. (1988). *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
- Bisol, L. (2010). O diminutivo e suas demandas. *DELTA*, 26 (1), pp. 59–85.
- Carvalho, H. de (1967). *Teoria da Linguagem. Natureza do Fenômeno Linguístico e Análise das Línguas*. Coimbra: Atlântida.
- Fandrych, I. (2008). Submorphemic elements in the formation of acronyms, blends and clippings. *Lexis – E-Journal in English Lexicology*, 2.
- Gonçalves, C. A. V. (2004). Usos morfológicos: os processos marginais de formação de palavras em português. *Gragoatá*, 21, pp. 219–242.
- Gonçalves, C. A. V. (2011). *Iniciação aos Estudos Morfológicos*. São Paulo: Contexto.
- Gonçalves, C. A. V. (2012). Atuais tendências em formação de palavras no português brasileiro. *Signum. Estudos de Linguagem*, 15, pp. 169–199.
- Lee, S. H. (1995). *Fonologia e Morfologia Lexical do Português*. Ph.D. dissertation, UNICAMP.
- Macambira, J. R. (1978). *Português Estrutural*, 2nd edn. São Paulo: Pioneira.
- Mattoso Câmara Jr., J. (1971). *Problemas de Lingüística Descritiva*. Petrópolis: Vozes.
- Rio-Torto, G. (1998). *Morfologia Derivacional. Teoria e Aplicação ao Português*. Porto: Porto Editora.
- Sandmann, A. J. (1989). *Morfologia Geral*. São Paulo: Contexto.
- Schwindt, L. C. (2000). *O Prefixo no Português Brasileiro: Análise Morfofonológica*. Ph.D. dissertation, Pontifícia Universidade Católica do Rio Grande do Sul.
- Villalva, A. (1994). *Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português*. Lisbon: FCG-FCT.
- Villalva, A. (2008). *Morfologia do Português*. Lisbon: Universidade Aberta.
- Villalva, A. (2009). Sobre a formação dos chamados diminutivos em Português Europeu. In *Actas do 25º Encontro da Associação Portuguesa de Linguística*, pp. 787–793.
- Villalva, A. and J. P. Silvestre (2014). *Introdução ao Estudo do Léxico. Descrição e Análise do Português*. Petrópolis: Vozes.
- Wetzels, W. L. (1992). Mid vowel neutralization in Brazilian Portuguese. *Cadernos de Estudos Lingüísticos*, 23, pp. 19–55.