

Compounding in Portuguese*

Alina Villalva

Portuguese compounds comprise a number of different structures, which explains the diversity of their morphological behavior and justifies the necessity of their identification and analysis. Furthermore, some structures, which are traditionally labelled as derived words, display many striking similarities with compounds. In this paper, I will try to set apart what is different and bring together what is comparable.

I will argue that the words usually labelled as compounds include root compounds (RtCs) and word compounds (WdCs). RtCs frequently display Latin or Greek borrowings as constituents and they are always head final, whereas WdCs, which are formed upon phrasal structures, may be right-headed or not. A correlation can be drawn between this distinction, motivated by the morphological category of the constituents, and Di Sciullo and Williams' (1987) proposals concerning compounds and syntactic words. Considering that the formation of WdCs (or syntactic words) corresponds to a native process (in contrast with the learned nature of RtC formation), I will present a typology of this kind of structures. I will also discuss the formation of syntactic words by reanalysis of a phrase as a word (cf. Di Sciullo and Williams 1987: 79). Inflection within syntactic words, their external inflection and evaluative suffixation, provide some evidence which supports the claim that, in some cases, reanalysis is optional, and, in some other cases, it is mandatory.

Finally, a section is dedicated to the discussion of the relation between derivation and compounding by looking into two specific structures: *mente* adverbs and evaluative words. Both of them display compound properties although they also share some of the features of derived words.

0. Introduction.

Compounding covers a wide range of different phenomena and raises difficult questions, such as those concerning the relationship between morphology, phonology and syntax, or historical change and synchronic productivity. These issues are particularly hard to deal with when ultra-

* I wish to thank Gabriela Matos and André Eliseu for their comments and suggestions that helped me to improve this work. I would also like to thank Alan Baxter and J. C. Clements for some editing advice. The remaining faults are, needless to say, my responsibility.

specialization in connection with continual theoretical progression shape contemporary work in linguistics.

As far as the description of Portuguese compounds is concerned, there is no reliable data source, the only available material is supplied by traditional grammars and by the knowledge of the language. This lack of interest may well be a consequence of the fact that, within this particular language, compounding is a much less "dependable" word formation process than derivation, but it provides no evidence that it is not an available device.

1. *Root compounds and word compounds.*

Forms such as those listed in (1) are traditionally classified as compounds, in Portuguese. However, the identification of the morphological category of their constituents and the analysis of their internal structure show us that they do not make up a homogeneous set.

- | | | |
|-----|-------------------------|--|
| (1) | biblioteca ¹ | (library) |
| | megalomania | (megalomania) |
| | guarda-jóias | (lit. keeps jewels = jewel case) |
| | fim de semana | (lit. end of week = week-end) |
| | belas-artes | (lit. beautiful arts = fine arts) |
| | amor-perfeito | (lit. love perfect = pansy) |
| | surdo-mudo | (deaf-mute) |
| | varapau | (lit. stick wood = shaft) |
| | vaivém | (lit. goes comes = fluctuation; shuttle) |
| | bomba-relógio | (lit. bomb watch = time-bomb) |

In fact, the interaction of several features causes a number of major distinctions. The first one is motivated by the morphological category of the constituents and it isolates those that are formed, at least, by one root² (cf. 2) from those which are exclusively formed by words (cf. 3):

- | | | |
|-----|--------------------------|----------------------------------|
| (2) | biblioteca | (library) |
| | herbívoro | (herbivorous) |
| | megalomania | (megalomania) |
| | piscicultura | (pisciculture) |
| | partidocracia | (excessive power of the parties) |
| | ministricida | (person who kills ministers) |
| | vasodilatação | (vaso-dilation) |
| | mundivisão | (perspective of the world) |
| | luso-brasileiro | (Luso-Brazilian) |
| | tragicômico ³ | (tragi-comic) |
| | cantautor | (singer-composer) |

¹ Graphic alternations such as the existence (versus the non-existence) of blanks or hyphens within these compounds are meaningless as far as their linguistic analysis is concerned.

² Scalise (1984: 76) identifies this type of constituents of Italian compounds as stems. I will consider the corresponding Portuguese elements to be roots, since they include no thematic indexes, nor inflectional affixes.

³ In this case, the first root (*tragic*) is probably reduced by haplology. Cf. *trágico-marítimo* "tragic (event) occurring in the sea".

- | | | |
|-----|-----------------|--|
| (3) | guarda-vestidos | (lit. keeps cloths = wardrobe) |
| | papa-formigas | (lit. eats ants = ant-eater) |
| | passatempo | (lit. spends time = pastime) |
| | fim de semana | (lit. end of week = week-end) |
| | pés-de-galinha | (lit. feet of chicken = crow's-feet) |
| | belas-artes | (lit. beautiful arts = fine arts) |
| | amor-perfeito | (lit. love perfect = pansy) |
| | pernalta | (lit. leg long = wader) |
| | surdo-mudo | (deaf-mute) |
| | actor-encenador | (actor-producer) |
| | vaivém | (lit. goes comes = fluctuation; shuttle) |
| | bomba-relógio | (lit. bomb watch = time-bomb) |
| | varapau | (lit. stick wood = shaft) |

1.1. *Root compounds.*

RtCs are particularly frequent in the domain of technical and scientific terminologies and in the formation of common 'savant' neologisms. Most of the roots that occur as constituents of RtCs are Latin or Greek borrowings, unavailable for other word formation processes, like derivation which usually selects the native corresponding forms:

- | | | | | |
|-----|------------------------------|------------------------------------|---|-------------------------------|
| (4) | <i>pisc</i> / <i>peix</i> : | <i>piscicultura</i> (pisciculture) | / | <i>peixaria</i> (fish market) |
| | <i>herb</i> / <i>er</i> : | <i>herbívoro</i> (herbivorous) | / | <i>peixeira</i> (fishmonger) |
| | <i>bibli</i> / <i>livr</i> : | <i>biblioteca</i> (library) | / | <i>eruanário</i> (herbalist) |
| | | | / | <i>livraria</i> (bookshop) |
| | | | / | <i>livreiro</i> (book seller) |
| | | | / | <i>grande</i> (big) |
| | | | / | <i>grandeza</i> (greatness) |

Some of these learned roots may occupy either the first or the second position in the structure of RtCs, but the head of the compound is always the rightmost constituent, as the corresponding paraphrases clearly demonstrate:

- | | | | | |
|-----|----|---|---|---|
| (5) | a. | [[<i>tecn</i>][<i>ol</i>] <i>cracia</i>] _{head} | = | power of the technicians (technocracy) |
| | | [[<i>pir</i>][<i>o</i>] <i>tecnie</i>] _{head} | = | technique related to fireworks (pyrotechny) |
| | b. | [[<i>graf</i>][<i>ol</i>] <i>mania</i>] _{head} | = | writing mania (graphomania) |
| | | [[<i>cal</i>][<i>ig</i>] <i>rafia</i>] _{head} | = | beautiful handwriting (calligraphy) |
| | c. | [[<i>morfol</i>][<i>og</i> ia]_ _{head} | = | knowledge of the form (morphology) |
| | | [[<i>poli</i>][<i>il</i>] <i>morfia</i>] _{head} | = | (occurrence of) many forms (polymorphism) |

The position of plural suffixes provides formal evidence for the identification of the head with the rightmost constituent (cf. *tecnocracias*, *pirotecnias*, *grafomanias*, *caligrafias*, *morfologias*, *polimorfias*).

There is a binding vowel (cf. Scalise 1984: 75-76, 99) between the two constituents of RtCs (frequently an *i* between roots of Latin origin and an *o* between roots of Greek origin). This binding vowel can also be found

between two native roots (eg. *luso-brasileiro*) or between a native root and a word: *verde-seco* 'lit. green dry' may occur as *verdesseco*, when the meaning is 'half-dry'.

The first constituent can be either a complement selected by the head or a modifier of the head:

- (6) a. [*tecn*]_{comp}[*eracia*]_{head}
 b. [*cal*]_{mod}[*grafia*]_{head}

Native roots can also combine with some learned roots (eg. [[*partid*]_{in}, *o* [*eracia*]_h]) or with other native roots or words (eg. [[*lus*]_o[*brasileiro*]]). The structure of semi-native RtCs is similar to the structure of learned RtCs (i.e. they are also right-headed and the leftmost constituent is either a complement or a modifier of the head), but strictly native RtCs, which are either nouns or adjectives, comprise yet another type. Structures belonging to this second type have a cumulative reading, which means that they assign conjointly two properties to the same referent. This makes it difficult to identify the head, from a semantic point of view, but inflectional marking occurs in a word final position, indicating that, in formal terms, they are also right-headed.

Finally, it is worth noting that RtC formation processes, which are processes of root concatenation, have some sort of counterpart in the WdC formation, which indicates that, in a way, this is the native equivalent process:

- (7) a. RtC: *herbívoro* = X who eats herbage (herbivorous)
 WdC: *papa-formigas* = W who eats ants (ant-eater)
 b. RtC: *biblioteca* = place where books are kept (library)
 WdC: *guarda-jóias* = place (object) where jewels are kept (jewel case)
 c. RtC: *vasodilatação* = dilation of blood-vessels (vaso-dilation)
 WdC: *fim de semana* = end of week (week-end)
 d. RtC: *caligrafia* = beautiful handwriting (calligraphy)
 WdC: *belas-arts* = beautiful arts (fine arts)
 e. RtC: *luso-brasileiro* = Luso-Brazilian
 WdC: *surdo-mudo* = deaf-mute

1.2. Word compounds.

WdCs are, nevertheless, very different from RtCs. One of the main questions concerns their identification. According to traditional grammarians, WdCs include only (i) an almost closed set of structures that do not have a compositional semantic interpretation (eg. *pés-de-galinha*, *amor-perfeito*), (ii) sequences of words which produce ungrammatical structures if they are phrasally analysed (eg. *guarda-vestidos* or *bomba-relógio*) and (iii) sequences that exhibit some kind of phonological merging (eg. *pernalta*).

The underlying criteria for this identification of WdCs are, thus, closely related to the existence of semantic amalgamation, or of some kind of syntactic "deviation".

One of the consequences of the adoption of these criteria is that structures like *pés-de-galinha* or *amor perfeito* are analysed as compounds, but similar structures, such as *selo de correio* "postage stamp" or *selo fiscal* "revenue stamp", are analysed as noun phrases. All these structures are syntactically transparent. Consequently, it is the occurrence of semantic drift that forces the first two to be treated as one lexical unit. The availability of two alternative analyses for similar structures⁴ is far from ideal, but there is another undesirable outcome which is related to the non-existence of a definition of semantic drift. In fact, native speakers hesitate in the classification of a great number of structures either as a WdC or as a phrase. I will try to demonstrate, in section 3, that this hesitation may be derived from the properties of these structures.

It is also worth looking into the phonological form of these words, and particularly into stress assignment. There is, in Portuguese, a phonological process of unstressed vowel raising, in both simple and derived words. For instance, when the vowel /o/ is unstressed, it is realized as [u] (cf. *bolo* ['bolu] "cake" / *bolimbo* [bu'liũu] "small cake").

WdCs, on the contrary, exhibit as many non-raised vowels as the number of constituent words, but there are some 'exceptions'. This is the case of a compound like *varapau*, which is formed by two words: *vara* "stick" and *pau* "wood". The first vowel of *vara* (when it occurs as a word) is a stressed vowel ([+ back, -round, + low]), but, when the word occurs as a constituent of *varapau* it becomes [-high, -low], which means that it is not a stressed vowel.⁵ Another example is *pernalta* "wader". In this case, not only is the first vowel of *perna* raised, but there is also an assimilation of its final vowel to the first vowel of *alta*.

In my opinion, words such as *varapau* and *pernalta* are no longer WdCs. Their compositional nature has been affected by diachronic changes which eliminated their internal structure and forced them to be lexicalized as simple words. The process may, in some cases, be seen in action: some native speakers pluralize sequences such as *cor de rosa* "rose-color, pink" or *cor de laranja* "orange-color" word-finally, just like simple and derived words (cf. *cor de rosas* [*sing*], *cor de laranjas* [*sing*]), contrarily to the 'grammatical' plural forms, which are *cores* [*sing*] *de rosa* and *cores* [*sing*] *de laranja*. Evaluative suffixation also corroborates the argument of an ongoing change. Both forms are accepted by native speakers, although the first one is preferred: *cor de rosinha* / *corzinha de rosa* "pale shade of pink".

⁴ Cf. [[*amor*]_N [*perfeito*]_{N,N} and [[*selo*]_N [[*fiscal*]_N]_{N,N}].

⁵ It has been brought to my attention that, for some speakers, the first vowel of *varapau* is a [+low] vowel. This clearly illustrates that lexicalization is a diachronic process, allowing the coexistence (at least for a certain period of time) of different outputs.

Finally, it should be noted that semantic drift does not trigger the operation of any phonological amalgamation processes (cf. *pés-de-galinha*). Traditional criteria are, thus, insufficient or inadequate for the identification of WdCs, but these shortcomings motivate other insights. Di Sciullo and Williams (1987)⁶ provide some useful tools for the analysis of compounds, which I will adopt here. In fact, they argue that compounds, like words formed by affixation, are head-final and syntactically opaque. Furthermore, they maintain that WdCs, like those occurring in Romance languages, are not compounds, but rather 'syntactic words', which constitute a class of objects that have a phrasal form (not a morphological one) but otherwise display the general properties of X⁰s.

Arguments in favour of the insertion of these words in X⁰ positions are well known (cf. Scalise 1984, DS&W 1987): no syntactic operation can insert, move or replace a constituent. Syntactic processes only have scope over the whole syntactic word:

- (8) a. **tina-nódoas de gordura*
(lit. removes spots of grease = remover of greasy spots)
- b. O *fim de semana* passou muito depressa.
(The week-end was soon over.)
**Que fim é que passou muito depressa?*
(Which end was soon over?)
*O *de semana*.
(The week one)
cf. O *que é que passou muito depressa?*
(What was soon over?)
O *fim de semana*.
(The week-end.)
- c. *tina-nódoas* → **tina-as*
(lit. removes spots → removes them = spot-remover → them-remover)

2. Types of syntactic words.

In the previous section, I suggested that DS&W (1987) analysis of Romance compounds as syntactic words is suitable for Portuguese WdCs. Consequently, the identification of possible syntactic words has to take into account their underlying phrasal structures. Since my purpose here is to present a morphological analysis, I will not go beyond elementary syntactic representations.

Syntactic structures involved in the formation of Portuguese syntactic words are APs, NPs and VPs. Structures such as those in (9) are also frequently considered to be compounds, but, in fact, they are derived words or at least closer to them than to compounds: although they are right-headed,

⁶ Hence DS&W (1987).

they are not RtCs, and although they are formed by two words, some of them can not have a well-formed underlying phrasal structure (cf. [*sempre*_{ADV} *viva*_{A,N}]_N or [*entre*_{prep} *olhar*_{V,IV}]).

- (9) a. [ADV A]_(A; N)
abaixo-assinado (undersigned; signed petition)
bem-intencionado (lit. well intended = well-minded)
mal-agrado (lit. badly thanked = ungrateful)
recém-nascido (lit. recently born = new-born baby)
sempre-viva (lit. always alive = houseleek)
- b. [P (A; N; V)]_(A; N; V)
antecipera (the day before the day before)
contra-revolucionário (counter-revolutionary)
entrelugar (lit. between look = to exchange glances)
sem-fim (lit. without end = immensity)
sem-número (lit. without number = countless number)
sem-sabor (lit. without taste = insipid)
sobre-excitação (overexcitement)
sobrescrito (lit. over written = envelope)
sobretudo (lit. over all = overcoat)

Phrasal structures underlying syntactic words are quite restricted. All de-AP syntactic words are formed upon coordinated structures (eg. *surdo-mudo* "deaf-mute"). De-NP words are typically formed upon an NP containing a head noun and an adjectival or nominal modifier, but they can also be formed upon a coordinated structure. They contain no specifiers and no pronominal reference. Finally, the underlying syntactic structure of de-VP words is typically a VP formed by a verb and its direct object, or two coordinated VPs.

2.1. Color name syntactic words.

The analysis of the underlying phrasal structures of syntactic words may sometimes be made difficult by words which are ambiguous in category. The color name terminology, for instance, provides a large number of syntactic words, which can surface both as nouns or adjectives, just like non-complex color names. This ambiguity is contextually annulled:

- (10) O *verde-esmeralda*_N é a minha cor favorita.
(Emerald-green is my favorite color.)
Comprei um lenço *verde-esmeralda*_A.
(I bought an emerald-green scarf.)
- (11) *Verde*_N é esperança
(Green means hope.)
*Olhos verdes*_A são traição.
(Green eyes are treason.)

There may also be a formal disambiguation when color name syntactic words are pluralized. Unlike simple color names, they display plural suffixes only when they have a nominal distribution:

- (12) *Os verdes*_[sing] -*esmeralda*_N são as *minhas cores preferidas*.
 **Os verde*_[sing] -*esmeralda*_N são as *minhas cores preferidas*.
*Comprei dois lenços verde*_[sing] -*esmeralda*_N.
 **Comprei dois lenços verdes*_[sing] -*esmeralda*_N.

Whenever possible, the category of color name words which are constituents of syntactic words must also be decided according to their internal structure. All of these syntactic words refer to shades of the color identified by the first constituent: *amarelo-canário* and *amarelo-pálido* are two different kinds of yellow, just like *azul-da-prússia* is a specific shade of blue.

- (13) a. *amarelo-canário* (lit. yellow canary = canary yellow)
branco-pérola (lit. white pearl = pearl-white)
verde-esmeralda (lit. green emerald = emerald-green)
vermelho-cereja (lit. red cherry = cherry color)
 b. *azul-da-prússia* (lit. blue of Prussia = Prussian blue)
 c. *amarelo pálido* (lit. yellow pale = pale yellow)
azul forte (lit. blue strong = royal blue)
branco-sujo (lit. white shabby = shabby white)
castanho-escuro (lit. brown dark = dark brown)
vermelho-vivo (lit. red alive = blood red)

In Portuguese, color adjectives always occur in a post-nominal position. Consider the following examples:

- (14) a. *Comprei um carro novo.* (I've bought a new car)
Comprei um novo carro.
 b. *Comprei um carro azul.* (I've bought a blue car)
 **Comprei um azul carro.*

This means that in (13a), where the second constituent is unquestionably a noun, the first constituent can not be an adjective, since it would be in a pre-nominal position. Furthermore, if it was an adjective, the meaning of *amarelo-canário* would be "a canary which is yellow" and this is an absolutely wrong paraphrase. In (13b), the PP (*da-prússia*) is a modifier of the first constituent, which is the head noun of the structure. Thus, and in order to reach a unified treatment, in (13c), although the internal structure does not provide any strong evidence supporting this analysis, the first constituent must also be a noun. Besides, the first constituent of these syntactic words can not undergo degree intensification, which is a characteristic process of morphological adjectival modification:

- (15) a. *Esta camisa é branquíssima.*
 (lit. This shirt is very white.)
 **Esta camisa é branquíssimo-pérola.*
 (lit. This shirt is very white pearl.)

- b. *Ela tinha uns olhos azulíssimos.*
 (lit. She had eyes very blue.)
 **Ela tinha uns olhos azulíssimos-claros.*
 (lit. She had eyes very blue light.)

There is, furthermore, a strong parallel between these three types of syntactic words that supports the analysis of the first constituent as a noun. On the one hand, a word like *azul marinho* "lit. blue sea_A = navy blue", which belongs to the third type, refers to a shade of the color that can also be given by a structure belonging to the first type (cf. *verde mar* "lit. green sea_N = sea green"). On the other hand, a word belonging to the third type, like *azul celeste* "lit. blue sky_A = sky blue", contains a constituent (*celeste*) which can be seen as a morphological counterpart of the PP *do céu*, a typical constituent of the second type.

2.2. [N N] syntactic words.

Another difficult case concerns those syntactic words that surface as [N N] sequences:

- (16) a. *actor-encenador* (actor-producer)
 b. *bomba-relógio* (lit. bomb watch = time-bomb)

(16a) has a cumulative reading: it refers to someone who is both an actor and a producer. (16b), on the contrary, does not refer to an object which is simultaneously a bomb and a watch, it identifies an object which is 'a bomb designed to explode at a pre-arranged time'. There is a formal argument that supports this distinction, and it is related to plural inflection. In the first case, plural suffixes attach to both constituents, whereas in the second one only the first constituent displays a plural marking:

- (17) a. *actores-encenadores*
 b. *bombas-relógio*

The first compound is formed upon a coordinated structure,⁷ whereas the second is formed upon a structure that contains a head noun and a nominal adjunct modifier. Some syntactic structures, such as the following which are delimited by square brackets, may be analysed as adjuncts to a projection of the head noun:

- (18) a. *O número [apresentado pelo domador de leões] é o melhor.*
 (The act performed by the tamer is the best one.)
 b. *Este leão [rei da jaula] está esfomeado.*
 (This lion king of the cage is starved.)

⁷ Some formally identical structures do not allow a cumulative reading, probably because the meaning has been lexicalized: a *tenente-coronel* is not a person who is both a *tenente* and a *coronel*. Nevertheless, I will consider that they belong to this type of syntactic words.

This second type of [N N] words also includes structures such as *mestre-escola* "schoolmaster", although the relation of the adjunct to the head noun is a different one.⁸

2.3. Inventory.

Thus, possible syntactic words, in Portuguese, may have the following underlying phrasal structures:

- (19) a. [A A]_{AP} *côncavo-convexo* (concavo-convex)
surdo-mudo (deaf-mute)
- b. [AP N]_{NP} *alta-roda* (lit. high wheel = high-life)
baixo-relevo (bas-relief)
boa-noite (good night)
curto-circuito (short circuit)
livre-pensador (freethinker)
má-língua (lit. bad tongue = slander)
meia-lua (half-moon)
novo-rico (nouveau rich)
quarta-feira (lit. fourth fair = wednesday)
verdes anos (lit. green years = youth)
- c. [N AP]_{NP} *água-furtada* (lit. water stolen = attic)
cofre-forte (lit. safe strong = strong-box)
estrela-cadente (lit. star falling = falling-star)
nódoa negra (lit. spot black = bruise)
pau-mandado (lit. stick ruled = servile person)
pé-direito (lit. foot right = height)
riso amarelo (lit. smile yellow = forced smile)
roupa branca (lit. clothing white = linen)
sangue-azul (lit. blood blue = blue blood)
selo-fiscal (lit. stamp tax = revenue stamp)
- d. [N N]_{NP} *andar-modelo* (lit. flat model = model-flat)
bomba-relógio (lit. bomb watch = time-bomb)
comboto-fantasma (lit. train ghost = mistery train)
couve-flor (lit. cabbage flower = cauliflower)
mestre-escola (lit. master school = schoolmaster)
mulher-polícia (lit. woman police = policewoman)
navio-escola (lit. ship school = training ship)
pexete-espada (lit. fish sword = sword-fish)
- e. [N PP]_{NP} *bola de neve* (lit. ball of snow = snowball)
cadeira de rodas (lit. chair of wheels = wheel chair)

⁸ This type of [N N] syntactic words is not very frequent in Portuguese, but in other Romance languages, like Italian (cf. *capostazione* "station master" or *fondo assistenza* "assistance fund", examples from Scalise 1984: 62, 125), it seems to be quite recurring.

- caminho de ferro* (lit. way of iron = railway)
fim de semana (lit. end of week = week-end)
lua de mel (lit. moon of honey = honeymoon)
moinho de café (lit. mill of coffee = coffee-mill)
moinho de vento (lit. mill of wind = windmill)
raio de acção (lit. ray of action = scope)
relógio de sol (lit. watch of sun = sundial)
selo de correio (lit. stamp of mail = postage-stamp)
- f. [N N]_{NP} *actor-encenador* (actor-producer)
comandante- chefe (lit. commander chief = commander-in-chief)
tenente-coronel (lieutenant-colonel)
- g. [V NP]_{VP} *abre-latas* (lit. opens tins = tin-opener)
afia-lápis (lit. sharpens-pencils = pencil-sharpener)
beija-mão (lit. kisses hand = hand-kissing)
faz tudo (lit. makes all = jack of all trades)
guarda-vento (lit. keeps wind = wind-screen)
limpa-chaminés (lit. cleans chimneys = chimney-sweeper)
quebra-mar (lit. breaks sea = sea-wall)
quebra-nozes (lit. breaks nuts = nutcracker)
tira-nódoas (lit. removes spots = spot-remover)
troca-tintas (lit. changes inks = bungler)
- h. [V V]_{VP}⁹ *cai-cai* (lit. falls falls = braceless bra)
chupa-chupa (lit. sucks sucks = lollipop)
come e dorme (lit. eats and sleeps = lazy-bones)
dói-dói (lit. hurts hurts = hurt)
sobe e desce (lit. goes up and comes down = merry go round)
vavém (lit. goes and comes = fluctuation)

3. Remarks on reanalysis.

In section 2, I have identified the underlying phrasal structure of Portuguese syntactic words. I will now try to demonstrate that that information is directly related to their morpho-syntactic behavior, namely in what concerns inflection.

In the domain of internal inflection and since syntactic words have a phrasal form, we expect agreement phenomena (gender and number) to be present inside the structure, whenever agreement requisites are met.

⁹ Under this heading I included two kinds of structures: the first one may be identified as a structure of intensification, by a kind of reduplication process (eg. *cai-cai*), whereas the second one is a structure of coordination of two different predicates. This distinction, however, does not determine a different behavior of the corresponding syntactic words.

This is, in fact, the case for syntactic words formed upon an NP containing an adjectival modifier, either in a pre- or in a post-nominal position:

- (20)
- | | | |
|----------------------|----------------------|------------------------------------|
| <i>boa</i> | <i>noite</i> | (good night) |
| $A_{[-fem, +sing]}$ | $N_{[+fem, +sing]}$ | |
| <i>verdes</i> | <i>anos</i> | (lit. green years = youth) |
| $A_{[-masc, -sing]}$ | $N_{[+masc, -sing]}$ | |
| <i>nódoa</i> | <i>negra</i> | (lit. spot black = bruise) |
| $N_{[+fem, +sing]}$ | $A_{[+fem, +sing]}$ | |
| <i>riso</i> | <i>amarelo</i> | (lit. smile yellow = forced smile) |
| $N_{[+masc, +sing]}$ | $A_{[+masc, +sing]}$ | |

Another case of internal inflection concerns the first type of de-VP syntactic words (cf. 19g). The first constituent is tense-marked. Not surprisingly, though, this verb form is a present tense, third person singular, which is, within the whole verbal paradigm, the less marked form, since it exhibits no person, number or tense morphemes, and it allows a generic interpretation. But there is a crucial difference between de-NPs like those in (20) and these de-VP words: the verb can not be further inflected, i.e. it is a frozen form:

- (21)
- | | |
|----|---|
| a. | * <i>abres</i> _[person] - <i>latas</i> |
| b. | * <i>abriu</i> _[past] - <i>latas</i> |

The second constituent, which is an internal argument of the verb, is very frequently a plural noun not required by agreement. This is probably due to the fact that this plural form favors a generic reading of the whole word (cf. *abre-latas* 'tin opener').

Some structures show us that the complement noun may also be a singular form:

- (22)
- | | |
|-------------------|---------------------------------------|
| <i>beija-mão</i> | (lit. kisses hand = hand-kissing) |
| <i>faz-tudo</i> | (lit. makes all = jack of all trades) |
| <i>quebra-mar</i> | (lit. breaks sea = sea-wall) |

It is possible to find some reasons for the occurrence of singular form nominal complements, either historical (*beija-mão* originally referred the ceremony of saluting the king or another equally important individual, whose hand was kissed), linguistic (*tudo* can not be inflected and it is intrinsically generic), or pragmatic (*quebra-mar* is a sea-wall that can not easily be moved from one location to another). But, crucially, these singular forms may occur because they also allow a generic reading.

The most significant data concerning the relevance of the underlying phrasal structure to the morpho-syntactic behavior of syntactic words is, however, provided by external plural inflection. We would expect inflectional processes to operate as within phrasal structures, but this is not always the case and it allows us to draw a second distinction between de-[+ N]Ps (which covers de-AP and de-NP words), on the one hand, and

de-VPs, on the other hand. The first group has a syntactic type of inflection, whereas the second exhibits a typical morphological inflection. This means that, in the first case, inflectional suffixes occur inside syntactic words, but in the second case inflection is word-final:

- (23)
- | | | | | |
|----|----------------------------|---------------|-------------------------|----------------------------|
| a. | $A_{[-sing]}$ | \rightarrow | $[surdos]_{AI[-sing]}$ | $AP_{[-sing]}$ |
| b. | $N_{[-sing]}$ | \rightarrow | $[moinhos]_{N[-sing]}$ | $PP_{NPI[-sing]}$ |
| | $N_{[-sing]}$ | \rightarrow | $[landares]_{N[-sing]}$ | $NPI_{[-sing]}$ |
| | $N_{[-sing]}$ | \rightarrow | $[livres]_{AP[-sing]}$ | $NPI_{[-sing]}$ |
| | $N_{[-sing]}$ | \rightarrow | $[nós]_{N[-sing]}$ | $NPI_{[-sing]}$ |
| | $N_{[-sing]}$ | \rightarrow | $[lactores]_{N[-sing]}$ | $NPI_{[-sing]}$ |
| c. | $[quebra]_{VI[+sing]}$ | VP | \rightarrow | $[quebra-mar]_{N[+sing]}$ |
| | $[quebra-mar]_{N[+sing]}$ | N | \rightarrow | $[es]_{[-sing]}$ |
| | $[chupa]_{VI[+sing]}$ | VP | \rightarrow | $[chupa-chupa]_{N[+sing]}$ |
| | $[chupa-chupa]_{N[+sing]}$ | N | \rightarrow | $[s]_{[-sing]}$ |

Evaluative suffixation operates in a similar way. De-[N]Ps exhibit an evaluative suffix near the head of the phrasal structure, whereas in de-VP words the evaluative suffix always occurs word-finally:

- (24)
- | | | |
|----|-------------------------|--|
| a. | <i>risinho amarelo</i> | (lit. smile-dim yellow = small forced smile) |
| | <i>peixinho-espada</i> | (lit. fish-dim sword = small sword-fish) |
| | <i>cadeira de rodas</i> | (lit. chair-dim of wheels = small wheel-chair) |
| b. | <i>abre-latazinho</i> | (lit. opens tin-dim = small tin opener) |
| | <i>dói-dóizinho</i> | (lit. hurt hurt-dim = small hurt) |

Notice that, if the evaluative suffix is associated to the first constituent of de-VP words, they become ungrammatical, but if the non-head of de-[+ N]Ps words exhibits that evaluative suffix, the sequence is accepted, although the meaning is different:

- (25)
- | | | |
|----|----------------------------|---------------------------|
| a. | * <i>abrinho-latas</i> | (lit. opens-dim tins) |
| | * <i>dóizinho-dói</i> | (lit. hurts-dim hurts) |
| b. | <i>riso amarelinho</i> | (a sort of forced smile) |
| | <i>cadeira de rodinhas</i> | (chair with small wheels) |

Furthermore, the distinction between de-VP and de-[+ N]P syntactic words is otherwise motivated. The lexical category of de-[+ N]P words is consistent with the category of their underlying phrasal structures. De-AP words are adjectives and de-NP words are nouns, regardless of their internal structure. The phrasal structure of de-NP words also allows us to calculate all morphological features of these syntactic words. Thus, it can be said that the morphological head coincides with the syntactic head (except when the

phrasal structure is a coordinated structure), and gender and number values coincide too:

- (26) A → [A A]_{AP}
 $N_{[\text{gender} = G, \text{number} = N]} \rightarrow [AP N_{[\text{gender} = G, \text{number} = N]}]_{NP}$
 $N_{[\text{gender} = G, \text{number} = N]} \rightarrow [N_{[\text{gender} = G, \text{number} = N]} AP]_{NP}$
 $N_{[\text{gender} = G, \text{number} = N]} \rightarrow [N_{[\text{gender} = G, \text{number} = N]} N]_{NP}$
 $N_{[\text{gender} = G, \text{number} = N]} \rightarrow [N_{[\text{gender} = G, \text{number} = N]} PPI]_{NP}$
 $N_{[\text{gender} = G, \text{number} = N]} \rightarrow [N_{[\text{gender} = G, \text{number} = N]} N]_{NP}$

The lexical category of de-VP syntactic words, on the contrary, is inconsistent with the category of its phrasal underlying structure which does not allow us to calculate the morphological structure of the syntactic word: the syntactic head of the phrasal structure (verb + inflection) can not be the morphological head, nor can it determine the gender or number of the compound. Nevertheless, all de-VP words are masculine, singular nouns, which are the unmarked values for gender and number. It may be argued, then, that, in this case, gender and number values are assigned 'by default' and that the morphological head is identified according to the unmarked position for morphological heads, which, in Portuguese, is the rightmost position:

- (27) a. $N_{[+ \text{masc.}, + \text{sing.}]} \rightarrow [V NP]_{VP}$
 b. $N_{[+ \text{masc.}, + \text{sing.}]} \rightarrow [V V]_{VP}$

According to DS&W (1987: 78-79), the operation that performs the mapping of a phrasal structure into an X^0 category is called reanalysis, which they define as a non-morphological word formation process, schematically represented as follows:

- (28) $X^0 \rightarrow YP$

Eliseu and Villalva (1991) suggest that reanalysis is a complex operation involving:

- (29) a. the neutralization of (some of) the properties of the internal structure of the underlying syntactic expression
 b. the derivation of a structure to which morphological processes can apply
 c. its structural and categorial reinterpretation

This is illustrated by the analysis of de-VP words (subtype 19g). It is argued that the properties of the node FLEX are neutralized by the fact that the verbal constituent can not be inflected (cf. 21). Thus, in this case, reanalysis operates by suppressing the head of the syntactic structure (FLEX) and deriving a [V N] sequence that will be structurally and categorially reinterpreted 'by default', according to general morphological principles. Furthermore, the thematic role assigned to the external argument of the

verb is reassigned to the syntactic word. In fact, these structures are subject-nouns (*afia-lápis* 'pencil sharpener' is a synonym of the derived word *afiador*, -dor being a subject-noun forming suffix), and their semantic interpretation is closely related to the interpretation of the underlying subject position:

- (30) agent noun: *limpa-chaminés* (chimney-sweeper)
 instrument noun: *afia-lápis* (pencil-sharpener)

The assignment of a [+N, -V] category to a VP (= V_{MAX}) is not easy to explain, but the existence of structures such as those in (30) shows us that, under given circumstances on which I will not elaborate here, sentences (IP, in the first case, and CP in the second one) and NPs have a similar distribution.

- (31) a. *Eles aprovarem a proposta será difícil.*
 (They to-approve-Agr the proposal will be difficult.)
 $[CP [N_{\text{max}} = IP \text{ eles } [I, [Agr]] \text{ aprovem a proposta}]]$
 $[I, \text{Inf}_2 \text{ ser difícil}]]]$
 (In Raposo 1987: 95)
- b. *Que eles aprovem a proposta é difícil.*
 (That they approve the proposal is difficult.)
 $[IP [N_{\text{max}} = CP [C, [C \text{ que}]] [IP \text{ eles } [I, \text{Inf}_1 \text{ aprovem a proposta}]]]]$
 $[I, \text{Inf}_2 \text{ ser difícil}]]]$

In the case of de-[+N]P structures, once the relevant syntactic properties have been neutralized, they are reinterpreted according to the morphological properties of the underlying syntactic head. But there are probably two different situations. The first one, which I will refer briefly although they require further investigation, concerns [A A] (cf. 19a) and [N N] (cf. 19d and 19f) sequences. Both (19a) and (19f) are coordinated structures involving X^0 categories. The following contrast seems to indicate that syntax excludes the coordination of lexical categories, which creates the necessary condition for the operation of reanalysis:

- (32) O rapaz e a rapariga
 (lit. The boy and the girl)
 *Os rapaz e rapariga
 (lit. The_{plural} boy and girl)

On the other hand, (19d) is a structure of adjunction of a lexical category (N) to a lexical category (N), also hypothetically excluded by syntax, which would again favor the operation of reanalysis. Notice that if the rightmost constituent was a PP or a relative clause the structure would be syntactically well-formed:

- (33) a. *mestre-escola*
 (lit. master school = schoolmaster)
mestre da escola
 (lit. master of the school)

- b. *navio-escola*
(lit. ship school = training ship)
navio que é uma escola
(lit. ship that is a school)

The second situation concerns the following structures:

- (34) [N AP]_{NP}
[AP N]_{NP}
[N PP]_{NP}

These structures can be sanctioned by syntax, unless a diachronic process of semantic or phonological amalgamation takes place and they are also the only cases that native speakers classify either as 'compounds' or as noun phrases (cf. section 1). In order to account for this contrast, it seems reasonable to posit that, in this case, reanalysis is optional, and that it is mandatory in all the other cases.

4. On the borderline between derivation and compounding.

Apparently, the distinction between derivation and compounding should not be problematic: derivation involves the selection of a base by an affix whereas compounding deals with sequences of roots or words. There are, however, some borderline cases, such as the formation of *-mente* adverbs and evaluative suffixation, which are two very common word formation processes in Portuguese, that should be taken into account.

4.1. *-mente* adverbs.

The adverb forming suffix *-mente* is very productive, although not completely unrestricted. It subcategorizes an adjective, and more precisely, its feminine form, unless the adjective is invariant:

- (35) *contínua* ADJ [+fem] → *continuamente* ADV
(continuous) (continuously)
contínuo ADJ [+masc] → **continuamente* ADV
(continuous) (continuously)
elegante ADJ [-masc, -fem] → *elegantemente* ADV
(elegant) (elegantly)

Since *contínua* is an adjective, *-a* is unquestionably an inflectional suffix that, once the adverb (*continuamente*) is formed, precedes *-mente*. If this adverb forming process is a derivational one, the well-known principle according to which derivation precedes inflection is, yet again, seriously threatened. It could be argued that some derivational processes may select

inflected words and *-mente* affixation might be one such process, but there are reasons to believe that this is not the case.

The inflectional suffix that occurs inside *-mente* adverbs behaves differently from those occurring in a word final position, since it is not relevant for syntax (gender is not even a pertinent category for adverbs). The gender for adjectives is decided by agreement, which means that the internal structure of *-mente* adverbs still recognizes *mente* as a feminine noun, which it historically was.

Furthermore, *-mente* adverbs display two stressed syllables, like syntactic words. The first one is the stressed syllable of the base adjective and the second is *-men-*. This means that neither of these syllables undergoes the phonological raising process that affects unstressed vowels.

These are arguments in favor of a treatment of *-mente* adverbs as syntactic words, formed upon an NP by mandatory reanalysis, although the head noun is a constant (like a suffix).

4.2. Evaluative words.

Evaluative suffixes can be augmentative, diminutive, affective, pejorative or, very frequently, a mix between some of these semantic types. For any given evaluative suffix there is another which is quite resembling: it is formed by the same phonetic sequence preceded by [z] (cf. *-ito/-zito*; *-ão/-zão*; *-eco/-zeco*). This second suffix always selects words, whereas the first one may select roots or words, under certain conditions:

- (36) *cas*_{NR} → *casinha*_N (house → little house)
*flor*_N → *florzinha*_N (flower → little flower)
*livr*_{NR} → *livrinho*_N (book → little book)
*livro*_N → *livrinho*_N (book → little book)

These suffixes can attach to nouns, adjectives or adverbs, and they never change or determine the syntactic category of the base, which distinguishes them from derivational suffixes:

- (37) *ded*_{NR} → *dedinho*_N (finger → little finger)
*ded*_{NR} → *dedão*_N (finger → big finger)
*ded*_{NR} → *dedeco*_N (finger → insignificant finger)
*nov*_{ADJR} → *novinho*_{ADJ} (new → very new)
*ced*_{ADV} → *cedinho*_{ADV} (early → very early)

In fact, evaluative suffixes seem to get category, gender and number values by internal agreement, and, from a semantic point of view, they are not the head of the forms where they occur. Like prefixes, they seem to behave as morphological modifiers. Their analysis as derived words is, thus, very problematic.

Let us now consider in particular the *z*-evaluative suffixes. As we can see in (38), an inflectional suffix (*-o*, *-a*) precedes the evaluative suffix.

In (39), we can furthermore observe that the selected base is plural:

- (38) *maci o* *zinh o* (very soft)
 R¹⁰ IS_(+masc) ES IS_(+masc)¹⁰
maci a *zinh a* (very soft)
 R IS_(+fem) ES IS_(+fem)
cão zinh o (small dog)
 N ES IS
 (39) *cão zinh o s* (small dogs)
 N ES IS IS

The inflectional suffixes that precede the z-evaluative suffixes in (38) and (39), however, behave again differently from those occurring in a word final position. Word final inflectional suffixes determine the inflectional categories that the word belongs to, but those that precede derivational suffixes do not and they are opaque to syntax.

Morpho-phonological evidence, again related to stress assignment, reinforces the parallel with *-mente* adverbs. The words where z-evaluative suffixes occur exhibit two stressed vowels, whereas those where the other evaluative suffixes occur do not:

- (40) *golinho:* [gu'li:nu] (small goal)
golozinho: ['golu'zi:nu] (small goal)

The main difference between *-mente* adverbs and z-evaluative words is that the former are right-headed whereas the latter are left-headed. This would not be a problem if z-evaluative words had a phrasal form, allowing us to analyse them as WdCs. But they have not. Consequently, z-evaluative words should be analysed as left-headed root compounds formed by adjunction of an evaluative suffix to a noun, adjective or adverb base. However, there are no left-headed root compounds, in Portuguese. Thus, z-evaluative words clearly illustrate an hybrid structure on the borderline between derivation and compounding.

5. Summary.

In this paper, I have drawn a distinction between root compounds and word compounds, based upon the morphological category of the constituents. RtCs, which have a morphological structure (eg. they are head final) are formed by a 'learned' process of root concatenation, whereas WdCs, which have an underlying phrasal form, are the output of native compounding. They both share the property of being syntactically opaque.

¹⁰ R = root
 IS = inflectional suffix
 ES = evaluative suffix

For the analysis of WdCs, I have adopted Di Sciullo and Williams' (1987) proposals concerning syntactic words, formed by reanalysis of a phrase as a word. However, inflection and evaluative suffixation show that reanalysis does not always operate in the same way.

Considering the underlying phrasal structure of syntactic words, we conclude (i) that reanalysis may either percolate their properties or assign them by default (eg. selecting the unmarked options for each particular category), and (ii) that it may either be optional or mandatory.

The relation between derivation and compounding has also been taken into consideration. *-mente* adverbs and evaluative words are borderline cases that basically apply compounding strategies to derivational structures.

Address of the Author:

Departamento de Linguística
 Faculdade de Letras - Universidade de Lisboa
 Alameda da Universidade
 P - 1699 Lisboa codex
 Portugal

References

- Di Sciullo, Anna Maria and E. Williams (1987), *On the Definition of Word*, Cambridge, Massachusetts, The MIT Press.
 Eliseu, André and Alina Villalva (1991), "Tira-teimas: entre morfologia e sintaxe", *Actas do 7º Encontro da Associação Portuguesa de Linguística* (116-140), Lisbon.
 Mateus, M. H. M., Amália Andrade, Maria do Céu Viana and Alina Villalva (1990), *Fonética, Fonologia e Morfologia do Português*, Lisbon, Universidade Aberta.
 Raposo, Eduardo (1987), "Case theory and Infl-to-Comp: the inflected infinitive in European Portuguese", *Linguistic Inquiry* 18: 85-109.
 Scalise, Sergio (1984), *Generative Morphology*, Dordrecht, Foris.
 Villalva, Alina (1986), *Análise Morfológica do Português*, Lisbon, Faculdade de Letras da Universidade de Lisboa.
 Villalva, Alina (1990), "Compounding in Portuguese", *ILTEC Working Papers* 3, Lisbon, Instituto de Linguística Teórica e Computacional.