

Compounding in contemporary Portuguese (to appear: *Probus* 2011)

GRAÇA RIO-TORTO*

SÍLVIA RIBEIRO**

Abstract

This article presents an overview of compounding in Portuguese. Conceived as a plurilexematic unit used as a holistic denomination, a compound is characterized by lexical atomicity. The continuum of compound classes that we propose is based on the morpho-lexical nature of the internal units (root, word) and on the (non)conformity of compound constructions with Portuguese syntactic templates. Since Portuguese compounds constitute a heterogeneous and borderline class, this analysis also focuses on the boundaries of compoundhood, namely those existing between compounds and phrasal nouns. This article also concentrates on the coordinative, subordinative, attributive and thematic relations between constituents and analyzes Portuguese compounds with respect to headedness in its morphological, semantic and categorical dimensions. Finally, we stress the existence of a narrow relationship between internal constituency, headedness and inflectional patterns of Portuguese compounds.

1. Introduction

In Portuguese a compound, i.e. a plurilexematic structure used as a holistic denomination and characterized by denotational unicity, consists of the combination of at least two lexemes – words or roots – into a larger lexical unit. Portuguese compounds are mainly nouns, whose internal constituents are related according to their semantic and grammatical features. These constituents can be grammatically linked by coordination, subordination or attribution (cf. Section 4).

The structure of this article is the following. Section 2 focuses on the nature of the constituents of compounds. Section 3 examines the delimitation of compounding regarding affixation and phrases. Section 4 describes the types of compounds according to their categorial and grammatical structure. Section 5 analyses Portuguese compounds with regards to headedness, endocentricity and exocentricity. Section 6 examines compounds' argument and semantic structure. Section 7 is devoted to the analysis of the patterns of inflection displayed by Portuguese compounds.

The Portuguese data are taken from Ribeiro (2010)¹ and from empirical data collected and described by Rio-Torto and Ribeiro for MorBoComp. Besides the

* CELGA, Universidade de Coimbra, Faculdade de Letras. gracart@gmail.com

**Universidade de Aveiro, ESTGA, e CELGA. silviasario@gmail.com We would like to thank the anonymous reviewers for their helpful comments and also Sandra Simões for the linguistic review.

¹ The corpus (3709 compounds) was collected in contemporary Portuguese dictionaries (*Dicionário da Língua Portuguesa Contemporânea*, da Academia das Ciências de Lisboa (2001) e *Dicionário Electrónico da Língua Portuguesa*, da Porto Editora (1996)) and in different press texts published along 2004-2006. **On-line data bases, such as /www.linguateca.pt/ have been consulted.**

descriptions of Portuguese compounds provided by the grammars of Cunha and Cintra (1998) and Mateus *et al.* (2003), we adopt the research framework used by Ribeiro 2010 to describe the main compound patterns of contemporary Portuguese. This study, grounded on a systematic description of a representative amount of data, is governed by different criteria applied in a consistent way. The theoretical background on compounding is inspired on Bisetto and Scalise 2005.

Some of the compounds we describe are lexicalized. Their inclusion in this work results from two main reasons: they have been constructed according to compounding patterns; the research we are doing about semantic lexicalization and the way these lexicalized structures are perceived is still in progress.

2. Basic units of compounds

Portuguese compounds may be based on different lexical constituents, namely words (*bilhete de identidade* (card of identity) ‘identity card’) and roots (*francófil(o/a)* (francophile) ‘francophile’).

2.1. Lexical categories

According to the lexical categories involved, as pointed out by previous works on Portuguese compounding (Baptista 1995; Cunha and Cintra 1998; Villalva 2000; Rio-Torto and Ribeiro 2009), there are different patterns of compounds. Some of them are specified on Table I.

Patterns of compounds	Examples
[VN] _N	<i>abre-latas</i> (open-cans) ‘can opener’
[NN] _N	<i>couve-flor</i> (cabbage-flower) ‘cauliflower’
[NA] _N	<i>era cristã</i> (era-Christian) ‘Christian era’
[AN] _N	<i>grande área</i> (big-area) ‘penalty area’
[NumN] _N	<i>mil-folhas</i> (thousand-leaves) ‘millefeuille’
[AA] _A	<i>surdo-mudo</i> (deaf-dumb) ‘deaf and dumb’
[[N[PrepN]]] _N	<i>boca de incêndio</i> (mouth of fire) ‘fire hydrant’
Neoclassical compounds ² :	<i>democracia</i> ‘democracy’

Table I: lexical categories involved in Portuguese compounding

In contemporary Portuguese the most productive patterns are [[N[PrepN]]]_N, [NA]_N, [VN]_N and [NN]_N, corresponding respectively to 44.1%, 37.9%, 11.4% and 6.6% of the data analyzed by Ribeiro (2010). The schemes [NV]_N, [[N[PrepV]]]_N, [VAdv]_N, [VV]_N, [VPron]_N, [AdvA]_A are residual.

In Portuguese, the [[N[PrepN]]]_N remains a very productive template of compounding (Baptista 1995). Structures such as **fim semana* (end week), **comboio passageiros* (train passengers) are unacceptable, because in Portuguese the presence of the preposition is still obligatory (*fim de semana* (end of the week) ‘weekend’, *comboio de passageiros* (train of passengers) ‘passengers train’). However, in contemporary Portuguese, as in Italian (cf. Bisetto and Scalise 1999), some [NN]_N compounds resulting from the loss of the preposition have started to appear (cf. Faria 2010). Some of them are terms with universal diffusion, such as *efeito estufa* (effect greenhouse) ‘greenhouse effect’, *seguro-saúde* (insurance health) ‘health insurance’,

² The roots involved in neoclassical compounds are also known as ‘semiwords’ (Scalise and Bisetto 2009)

poupança-reforma (saving retirement) ‘retirement savings’.

2.2. Morphological structure

The constituents used as inputs for Portuguese compounds are mainly morphologically simple. However, some compounds (cf. Table II) present a suffixed form, denominal or deverbal (Ribeiro 2010).

one suffixed constituent	denominal A	[NA] _N	<i>bomba atômica</i> (bomb atomic) ‘atomic bomb’
	deverbal N	[[N[PrepN]]] _N	<i>travão de mão</i> (brake of hand) ‘hand brake’
two suffixed constituents	deverbal N + denominal A	[NA] _N	<i>planeamento familiar</i> (planning family) ‘family planning’

Table II: Suffixed input units

Contrarily to what occurs in Russian, for instance, where suffixes create “compound-suffixed words” (Benigni and Masini 2009: 175), in Portuguese suffixation does not function as an active procedure to create new compounds. In fact, Portuguese compounding includes suffixed units already existent, so that compounding is not characterized by the insertion of a suffix specially used for that purpose. Nevertheless, derivational processes may apply to a morphological compound: in these cases, the scope of suffixation is the compound base ([[*ginec*]+*o*+[*log*]]+*ista*) (gynecologist) ‘gynecologist’).

2.3. Morphological, morphosyntactic and phrasal or syntactic compounds

Portuguese compounds are divided into three groups: morphological compounds, morphosyntactic compounds and phrasal or syntactic compounds (cf. Rio-Torto and Ribeiro 2009)³. The *continuum* of these classes is based on the morpho-lexical nature of the internal units (root, word) and on the (non)conformity of compound constructions with Portuguese syntactic templates.

Morphological compounds include at least one root, frequently neoclassical, and their constituents are joined by a linking vowel. The linking vowel is *-i-* (*aquífer(o,a)* ‘aquifer’) typically when the second constituent is [+Linate] and *-o-* in all other

³ Following Di Sciullo and Williams (1987), Villalva (2003) displays a different characterization of Portuguese compounds: morphological and (morpho)syntactic compounds are root and word based, respectively, as we also claim here. But categorial and semantic patterns proposed by the author exclude compounds [[N[PrepN]]]_N, a very representative pattern of phrasal compounds in Portuguese. Moreover, [N[prepN]]_N and [NA]_N compounds whose reading is not compositional are seen by the author as “lexicalized syntactic expressions”. Contrarily, we claim that lexicalization is transversal to all classes of complex words, so it is not a definitory dimension of a compound’s class. The criteria for the triple distinction we propose are the morpho-lexical nature of the constituents, in articulation with their syntactic behavior. Headedness, ignored as a crucial domain in previous works on Portuguese compounds, is here deeply described in its morphological, categorial and semantic dimensions.

situations *bibliófil(o,a)* ‘bibliophile’). However, these regularities are not absolute: *gasoducto* and *oleoducto* ‘gas and oil pipe respectively’ display a linking vowel *-o-* and *-ducto* is [+Latinate].

Morphological neoclassical compounds, such as *aquífer(o,a)* ‘aquifer’ or *bibliófilo* ‘bibliophile’ correspond to one stress domain. Other morphological compounds, such as *ibero-americano* ‘iberoamerican’ or *físico-química* ‘physics-chemistry’ correspond to two prosodic words (Vigário 2003) (cf. Section 3.2.1). With morphological compounds inflection only occurs at the end of the compound (*luso-americano(s)*, ‘*luso-americans*’) having scope over the whole. The absence of internal inflection proves that morphological compounds combine roots, not words.

A morphosyntactic compound (Table III), interpreted as resulting from a process of univerbation of a syntactic structure, involves the presence of at least two words. The structure of the final construction does not obey the syntactic rules governing internal phrasal structure (cf. Brito 2003).

[NN] _N	<i>decreto-lei</i> (decree-law) ‘decree law’
[AA] _A	<i>morto-vivo</i> (dead-alive) ‘living dead’
[VV] _N	<i>pára-arranca</i> (stops-starts) ‘stop-starts’

Table III: Morphosyntactic compounds

Actually, when used in a phrasal structure, these words are not allowed to co-occur without the presence of a conjunction (*e* ‘and’), as the ungrammatical label of 1a) illustrates.

1) *escritor-actor* (writer-actor) ‘writer-actor’:

a) **O escritor actor da tragédia é muito conhecido.* ‘The tragedy’s **writer actor** is well known.’

b) *O escritor e actor da tragédia é muito conhecido.* ‘The tragedy’s **writer and actor** is very well known’

Since these word combinations do not respect the syntactic configuration of phrases, they cannot be used as grammatically acceptable arguments of a propositional structure. Univerbation of some [NN]_N compounds is attested by lexical gender of the whole, which is governed by head gender, as in *decreto-lei* (decree law) ‘decree law’, *jardim-escola* (garden school) ‘kindergarten’:

2) [+Masc][_O[+Masc] [*decreto*_[+Masc] *lei*_[-Masc]]]

3) [+Masc][_O[+Masc] [*jardim*_[+Masc] *escola*_[-Masc]]]

As data collected by Ribeiro (2010) emphasize, some VN compounds also display the characteristic properties of morphosyntactic compounds once they do not obey syntactic rules. Indeed, sequences as those in (4) do not follow the typical internal structure of verbal phrases in Portuguese, as they exhibit as verbal complement a singular countable noun (*mão, mar, bicho, língua*) not preceded by an article. These VN structures depart from typical syntactic structures where a determiner is mandatory if a singular, non massive noun is present (5-6).

4) [VN]_N: *beija-mão* (kiss hand) ‘hand-kissing’, *corta-mar* (cut sea) ‘break-water’, *mata-bicho* (kill bug) ‘tot’, *trava-língua* (brake tongue) ‘tongue twister’

- 5) *O muro que **corta o** mar é resistente*
(the wall that cuts the sea is tough)
- 6) **O muro que **corta mar** é resistente*
(the wall that cuts sea is tough)

Syntactic or phrasal compounds (Table IV) involve word sequences whose internal structure obeys the syntax rules typical of phrases.

[[N[PrepN]]] _N	<i>caminho de ferro</i> (way of iron) ‘railway’
[[N[PrepV]]] _N	<i>ferro de engomar</i> (iron of starch) ‘flat iron’
[VconjV] _N	<i>come e dorme</i> (eat and sleep) ‘eat and sleep’
[NA] _N	<i>era cristã</i> (era Christian) ‘Christian era’
[AN] _N	<i>belas-artes</i> (beautiful arts) ‘fine arts’

Table IV: syntactic or phrasal compounds

Some [VN]_N compounds (7), which include a pluralized noun functioning as verbal complement, obey syntactic rules typical of verbal phrases (cf. 8)

- 7) [VN]_N: *abre-latas* *guarda-jóias*
(open cans) (keep jewels)
‘can opener’ ‘jewel box’

8)

a) *Ele **abre latas** com facilidade*
(He opens cans easily)

b) *Ele **guarda jóias** raras em casa*
(He keeps rare jewels at home)

The most representative subtypes of syntactic compounds are [[N[PrepN]]]_N and [NA]_N compounds (Ribeiro 2010). Despite the similarities between phrases and phrasal compounds, these ones present a fixed configuration — anchored on a maximal opacity to lexical internal variation —, in contrast with the greatest possibility of variation of phrases (cf. Section 3.2.2). In Romance languages, as phrasal compounds whose head is a N admit internal pluralisation, inflexion is not a crucial criteria for compounding.

3. Delimitation and frontiers of compounds

3.1. Compounding and derivation

In Portuguese, compounding consists of the combination of two lexemes, words or roots. Differently, derivation is characterized by the addition of an affix to a lexeme — a root or a stem.

Affixes (here underlined) have a fixed position inside the word (*inovação* ‘innovation’ and *desintegrar* ‘disintegrate’, but **ção+inova*, **integra+des* and they cannot combine with each other to form a new word (**des+ismo*, **in+ção*,

**re+mento*)⁴. Contrarily, in compounding, the combination is more flexible, even when the constituents are roots. Some neoclassical roots are initial combining forms (*caligrafia* ‘calligraphy’), others final combining forms (*insecticida* ‘insecticide’), and some others can occur in both positions (*filantropia* ‘philanthropy’/*cinéfilo* ‘cinophile’, *pedicure* ‘pedicure’/*ortopedia* ‘orthopedics’). Some [-learned] words also exhibit a flexible position inside compounds: e.g. *disco duro* ‘hard-disk’/*cheque-disco* ‘check-disk’; *escola-piloto* ‘pilot-school’/*navio-escola* ‘training ship’. Nevertheless compounding and derivation deeply interact (Stekauer 2005, Ralli 2010) and so it is not an easy task to clearly define the borderlines between both types of complex words. Rio-Torto (2006) explains that Portuguese adverbs with *-mente* and z-evaluative suffixed words are frontier cases, since they exhibit characteristic from derived words and compounds. For instance, z-evaluative suffixed words not only demand external pluralization, typical of derived words (*olhitos* ‘little eye(s)’), but also internal pluralization (*lençoizinhos* (little sheets), *pãezinhos* (little breads), as frequently happens with compounds (*lençóis de banho* (sheets of bath) ‘bath sheets’, *pães de leite* (breads of milk) ‘milk breads’). Gonçalves (forthcoming) considers that some *splinters* (*-graf-*, *-log-*, *-metr-*) are in a process of grammaticalisation and semantic specialization (*-(ó)log-* ‘specialist’, *-(ó)graf-* ‘instrument, specialist’, *-(ó)metr-* ‘unit of measure, instrument), behaving as suffixe-like formants.

3.2. Compounds and phrases

A source of complex words is univerbation of phrases which function as a single lexical unit in the mental lexicon, as the following examples illustrate.

- | | |
|-----------------------------|-----------------------|
| 9) <i>cartão de crédito</i> | <i>olhos de lince</i> |
| (card of credit) | (eyes of lynx) |
| ‘credit card’ | ‘eagle yes’ |

The inputs of these compounds are similar to those of phrasal structures: in fact, the [N[PrepN]] configuration may correspond to a compound [[N[PrepN]]_N] or to a noun phrase of the kind [N[PrepN]]_{NP}. Different criteria — prosodic, semantic, morphological, lexical-syntactic — may be invoked to distinguish compounds from phrases. The next sections are devoted to their analysis.

3.2.1. Prosodic structure

In Portuguese, stress position is not a useful criterion to distinguish a [NA/N] or [[N[PrepN]]] compound from a phrase. Differently, in Germanic languages, stress position distinguishes compounds from phrases: compound stress falls on the non head constituent (Eng. ‘blackboard’, Dutch ‘rotkohl’) and phrasal stress occurs on the noun phrase head (Eng. *black ‘board*, German *Rode ‘Kool*). This criterion is not applicable to Portuguese, because in Portuguese compound stress always falls on its last phonological word (Vigário 2003, 2010). Some morphological compounds are a single phonological word (*hortícola* ‘horticultural’, *psicólogo* ‘psychologist’); other include two phonological words (*astro-física* ‘Astrophysics’, **foto-térmico** ‘photothermal’ (the stressed syllable is in bold).

⁴ For further information about the boundaries between prefixes and compounds see Rio-Torto (in print).

3.2.2. Compounds' syntactic and lexical opacity

Lexical integrity and syntactic atomicity of compounds are two major features of compoundhood (Lieber and Scalise 2007): neither deletion nor lexical modification of constituents is allowed in the interior of compounds.

Neither of these two criteria are controversial with morphological compounds nor with [VN]_N compounds. The problems emerge with [[N[PrepN]] and with [NA]_N structures, because they are on the face similar to noun phrases, and the boundaries between both structures remain a serious issue to linguistic theories.

The following five tests (Bisetto and Scalise 2005, Lieber and Scalise 2007) support the delimitation between compounds and phrases: (i) head deletion under coordination, (ii) wh-movement, (iii) topicalisation, (iv) pronominal reference and (v) insertion of lexical material. A positive answer to these tests proves the phrasal nature of the expression and a negative result means that the structures are compounds. The Portuguese compounds *base de dados* (base of data) 'database', *guarda-costas* (save back) 'bodyguard', *hidroavião* (hydroplane) 'hydroplane', *linhas aéreas* (lines aerial) 'airlines' do not consent any of the mentioned tests.

The internal opacity to any lexical changing remains the main test of compoundhood in Portuguese, as lexical insertion between the constituents or any other modification is normally blocked, as examples in (10-12) illustrate:

10)

- a) *moinho de vento*
(mill of wind)
'windmill'
- b) *moinho *antigo de vento*
(mill *ancient of wind')
- c) *moinho de *muito vento*
(mill of *much wind)

11)

- a) *saca-rolhas*
(pull corks)
'corkscrew'
- b) *saca-*facilmente-rolhas*
(screw *easily corks)

12)

- a) *vinho branco*
(wine white)
'white wine'
- b) *vinho *completamente branco*
(wine *completely white)
- c) *vinho *branquíssimo*
(wine *very white)

Also the adjectivation strategies reveal the special nature of phrasal compounds: in compounds, adjectivation, when possible, does not agree with any of the internal constituents of the compound and its scope doesn't impact individually on the internal terms, emphasizing the structure's internal opacity.

13)

- a) *o lança-chamas perigoso* or *o perigoso lança-chamas*
(the dangerous flame-thrower)

- b) *lança-chamas *perigosas* or *lança-*perigosas chamas*
 ‘throw dangerous flames’

Constructions such as (13a), where the adjective *perigoso* ‘dangerous’, in the masculine form, predicates the whole compound, are grammatical. Contrarily, the constructions (13b) are ungrammatical, because the adjective *perigosas*_[-Masc] modifies only one of the constituents of the compound, as proved by the gender agreement with the feminine noun *chamas*.

Also the reaction of [[N[PrepN]]_N] compounds with respect to adjectivation confirms the internal opacity of Portuguese compounds. Indeed, [[N[PrepN]]_N] internal opacity is supported (i) by the gender of the modifying adjectives and (ii) by their scope. Let us take into account the Portuguese equivalent to English ‘identity card’ — *bilhete de identidade* —, with gender labels (14) and modified by the adjective *português* (15).

- 14) _[+Masc][_o_[+Masc] [*bilhete*_[+Masc] *de identidade*_[-Masc]]]
 (the card of identity)
 ‘identity card’
- 15) _[+Masc][[_o[*bilhete*_[+Masc] *de identidade*_[-Masc]] *português*_[+Masc]]
 (the card of identity Portuguese)
 ‘the portuguese identity card’
- 16) _[+Masc][*o bilhete de identidade *portuguesa*_[-Masc]]
 (the card of identity Portuguese)
 ‘*the card of Portuguese identity’

The structure in (15), in which the scope of the adjective is the whole compound (*bilhete de identidade*) and the adjective’s masculine form (*português*) is determined by the gender of the noun head (*bilhete*), is acceptable. Contrarily, the feminine adjectival form **portuguesa* (16) is unacceptable, showing that adjectival modification of part of the compound (*identidade*, a feminine noun) is impossible.

In Portuguese compounds, changes in the conditions of determination are also blocked, as they could completely modify the meaning and the reference, mainly when the structure is [[N[PrepN]]_N]. The compound [*fim*_[+Masc] *de semana*_[-Masc]]_{N[+Masc]} (end of week) ‘weekend’ refers to the period of rest that goes from the end of the working week until the end of Sunday. In *fim da semana* (end of the week) the presence of the feminine definite article *a* (*de+a>da*) means that the *semana* refers to a particular week, determined (con)textually or otherwise located. When definite determination is activated, the [[N[PrepN]] structure *fim da semana* (‘end of **the** week’) refers to the last working days of a specific week. Thus, when a prepositional phrase is a compound’s constituent, no internal changing of the determination is allowed. Any change in the determination conditions modifies the meaning of the whole structure and has direct influence on its lexical status.

The (im)possibility of lexical insertion, the degree of (un)determination and the scope of adjectivation are assumed as important tests of compoundhood, because of the referential effects they produce and the opacity of the whole they reveal. Particularly when analyzing [[N[PrepN]]_N] structures, we believe that if internal changing is allowed, we are no longer dealing with compounds ([[N[PrepN]]_N]) but with noun phrases ([[N[PrepN]]_{NP}]).

4. Types of compounds

4.1. *Categorial classes of compounds*

As in other languages (Guevara and Scalise 2009), in Portuguese compounding the predominant output category is Noun. Adjectives and verbs are less frequent and adverbs are residual.

In Portuguese, the compound's lexical category is not always identical to the category of the constituents: in some cases, the lexical category of the output coincides with that of the constituents ($[\text{NN}]_N$, $[\text{AA}]_A$), but in other situations there is no equivalence. In these cases ($[\text{NA}]_N$, $[\text{AN}]_N$, $[\text{NumN}]_N$), the lexical category of the compound is determined by the lexical category of its head. Portuguese, as well as Italian and Chinese, have 'Absolute Categorial Exocentricity' (Scalise, Fabregas and Forza 2009), that is, there are structures ($[\text{VV}]_N$) whose output category is completely different from the input categories.

When none of the constituents transmits their category to the compound, i.e. when no categorial head exists and the compound is exocentric, the lexical category and the gender of the whole emerge in syntax since a determiner is present ($[\text{VV}]_N$ *o vaivém* (go come) 'space craft'; $[\text{VAdv}]_N$ *o/a fala-barato* (speak cheap) 'gasbag').

4.2. *Internal grammatical relations: coordinate, subordinate and attributive compounds*

Compounds in Portuguese can be divided into three classes determined by the type of grammatical relations established between their constituents: coordinate compounds, subordinate compounds and attributive compounds (Bisetto and Scalise 2005; Rio-Torto and Ribeiro 2009; Ribeiro 2010).

Coordinate compounds are based on a symmetrical relation between the constituents. Subordinate compounds are based on a predicator-argument completeness and attributive compounds imply a relationship of modification between their constituents.

4.2.1. *Coordinate Compounds* The constituents of coordinate compounds belong obligatorily to the same lexical category and are linked by a copulative operator, either explicit (*sal e pimenta* (salt and pepper) 'salt-and-pepper') or not (*pateta-alegre* (fool happy) 'happy fool'). The relationship between the two constituents is symmetrical and additive.

Two subclasses can be identified inside this group: a) appositional compounds, such as *poeta-pintor* (poet-painter) 'poet-painter', where the whole represents a hyponym of both constituents; and b) copulative compounds, such as *Alsace-Lorraine* (regions), *norte-sul* 'north-south'. In these cases the compound is not a hyponym of either constituent. This distinction has visible reflexes in pluralization (cf. Section 7).

The templates of Portuguese coordinate compounds are those on Table V.

$[\text{NN}]_N$	<i>norte-sul</i> 'north-south', <i>poeta-pintor</i> (poet-painter) 'poet-painter'
$[\text{AA}]_A$	<i>lusofrancês</i> 'lusofrench', <i>surdo-mudo</i> (deaf-dumb) 'deaf and dumb'
$[\text{VV}]_N$	<i>levar-traz</i> (take-bring) 'gossip', <i>salva-guarda</i> (save-keep) 'safeguard'

Table V: Portuguese coordinate compounds

4.2.2. *Subordinate compounds* In subordinate compounds the two constituents are connected by a predicator-argument relation, i.e. a relation of implicit dependence,

namely a subordinative one, as in *guarda-costas* (save-back) ‘bodyguard’. The non-head (*costas* ‘back’) is a complement (the Theme) of the (de)verbal head (*guarda* ‘save’).

In Portuguese subordinate compounds, the head (here underlined>) is mainly on the left (*condutor de taxi* (driver of taxi) ‘taxi driver’). Nevertheless, neoclassical compounds (*insecticida* insecticide) ‘insecticide’) present the head on the right. [NV]_N compounds are atypical, because the internal argument is on the left. The noun *sanguessuga* (blood suck) ‘leech’ is probably no longer perceived as a compound.

Table VI presents the templates of Portuguese subordinate completive compounds.

[VN] _N	<i>lava-louça</i> (wash tablewear) ‘dishwasher’
[[N[PrepN]]] _N	<i>processador de texto</i> (processor of text) ‘word processor’
[NA] _N	<i>planeamento familiar</i> (planning family) ‘family planning’
[VPron] _N	<i>faz-tudo</i> (do everything) ‘handyman/clown’
Neoclassical roots	<i>cancerígeno</i> ‘carcinogen’

Table VI: Portuguese completive compounds

The predicate can be a verb form ([VN/Pron]_N, [NV]_N) or a deverbal noun ([NA]_N). In [[N[PrepN]]]_N structures the head is a deverbal noun and the prepositional phrase [PrepN] functions as its Theme argument (cf. Section 6). With neoclassical compounds, the predicate is on the right and the left constituent is the internal argument.

4.2.3. *Attributive compounds* In attributive compounds, the compound’s constituents are linked by a modified-modifier relation. The head constituent, mainly on the left, is specified and/or modified by the non-head constituent, which, restricting the reference of the head, does not function as a selected argument. Despite the large majority of attributive compounds have the head on the left, there are some attributive compounds whose head is on the right: [A/NumN]_N or [AdvA]_A.

The non-head of an attributive compound, representing a PROPERTY that specifies an attribute of the head, can have a (i) classificatory or taxonomic function, and the whole denotes a subclass of the class codified by the head, or a (ii) qualifying and/or evaluating function, attributing non defining properties (shape, colour, quality) to the head.

Modification/attribution may be expressed by an adjective, a noun, a Prepositional Phrase, an adverb, a numeral, as illustrated on Table VII.

[NA] _N	<i>arroz doce</i> (rice sweet) ‘rice pudding’ <i>horário nobre</i> (schedule noble) ‘prime time’
[NN] _N	<i>cartão-jovem</i> (card young) ‘youth card’ <i>palavra-chave</i> (word key) ‘key-word’
[[N[PrepN]]] _N	<i>chapéu-de-chuva</i> (hat of rain) ‘umbrella’ <i>homem de palavra</i> (man of word) ‘man of his word’
[NumN] _N	<i>primeiro-ministro</i> ‘prime minister’ <i>terceiro mundo</i> ‘Third World’
[AN] _N	<i>alto-relevo</i> ‘high relief’ <i>livre arbítrio</i> ‘free will’

[AdvA] _A	<i>bem-humorado</i> ‘good humoured’ <i>malcontente</i> ‘malcontent’
---------------------	--

Table VII: Portuguese modificative compounds

Portuguese data (Ribeiro 2010) show that attributive compounds are prevalent (≈82%), followed by subordinate (≈14%) and coordinate compounds (≈4%). This order (ATT>SUB>COORD) departs Portuguese from what seems to be the general tendency (SUB>ATT>COORD) identified by Guevara and Scalise (2009) and reflects the predominance of [NA]_N and [[N[PrepN]]_N patterns which are mainly based on attributive relations between their constituents⁵.

4.3. Transversal grammatical and semantic relations

Some grammatical relations underlying compounds are transversal to different patterns of compounding: [NA]_N, [[N[PrepN]]_N and [NN]_N patterns are the most multivalent, as they can be based on different grammatical relations between their constituents (cf. Rio-Torto and Ribeiro 2009; Ribeiro 2010). The patterns containing a Verb support mainly internal subordinate relations. Table VIII schematizes the pervasiveness of these internal relations.

	Coordinative relations	Subordinative relations	Attributive relations
[[N[PrepN]] _N , [NA] _N	-	+	+
[AN] _N	-	-	+
[NN] _N	+	-	+
[VN] _N ⁶	-	+	-

Table VIII: Transversal internal grammatical relations in Portuguese compounds

Coordinate appositional compounds (cf. Section 4.2.1), denoting a hyponym of both constituents, are mainly used as human denominations (*trabalhador-estudante* (worker student) ‘student-worker’), or as designations for commercial establishments (*bar-restaurante* (bar restaurant) ‘bar-restaurant’) and for events (*jantar-comício* (dinner rally) ‘dinner-rally’). Copulative compounds denote a relationship between the two constituents (*mãe-filho* (mother son) ‘mother-son’).

Subordinate completive compounds (cf. Section 4.2.2) are based on a predicator-argument relation: the constituent on the right functions as the obligatory complement – the Theme – of the predicate (the verbal or deverbal constituent on the left). They denote mainly instruments (*abre-latas* (open cans) ‘can opener’), agents (*limpa-chaminés* (clean chimneys) ‘chimney sweep’), plants (*girassol* (turn sun) ‘sunflower’) or animals (*beija-flor* (kiss flower) ‘hummingbird’).

The non-head constituents of attributive compounds (cf. Section 4.2.3) have ontological and semantic restrictive and qualifying functions, involving various kinds of relations — those of ownership, whole/part, location, instrumentality —, depending on the referential nature of what is denoted (cf. 17).

17)

⁵ In fact, subordinative relations are not very common with these two patterns, occurring mainly when the N on the left is deverbal.

⁶ [NV]_N neoclassical compounds such as *cancerígeno* ‘carcinogen’, *fungicida* ‘fungicide’ behave as subordinative compounds.

- a. *cadeira de rodas* (chair of wheels) ‘wheelchair’: *rodas* ‘wheel’ denotes a **part** of the chair;
- b. *hidroginástica* (hydrogymnastics) ‘hydro-gymnastics’: *hidr-* ‘hydro’ denotes the **locative**;
- c. *cartão jovem* (card young) ‘youth card’: the noun *jovem* denotes the **beneficiary**;
- d. *nevralgia* ‘neuralgia’: *nevr-* ‘neuro’ denotes the **path** of the pain;
- e. *biocombustível* ‘biofuel’, *carro a gasóleo* (car (powered with) diesel) ‘diesel car’: *bio-* and *gasóleo* represent the **raw material** of combustion.

Portuguese compounds are frequently used as designations for concrete entities that belong to lexical-conceptual areas such as those of animals, instruments, and human denominations (cf. Table IX). Compounds are not available to form abstract nouns: those are mainly formed by suffixation: *-ismo* (*minimalismo* ‘minimalism’) or *-idade* (*ruralidade* ‘rurality’).

	[VN] _N	[NN] _N	[NA] _N	[[N[PrepN]]] _N
Animals	<i>beija-flor</i> (kiss flower) ‘hummingbird’	<i>peixe-espada</i> (fish sword) ‘swordfish’	<i>cavalo-marinho</i> (horse marine) ‘sea horse’	<i>bicho-da-seda</i> (bug from silk) ‘silkworm’
Instruments	<i>corta-relva</i> (cut grass) ‘lawnmower’	<i>lâmpada-piloto</i> ‘pilot-lamp’	<i>arma branca</i> ‘sharp weapon’	<i>tubo de ensaio</i> ‘test tube’
Humans ⁷	<i>limpa-chaminés</i> (clean chimneys) ‘chimney sweep’	<i>piloto-aviador</i> (pilot-aviator) ‘airman’	<i>defesa-direito</i> (defender right) ‘right back’	<i>pé-de-chumbo</i> (foot of lead) ‘slow coach’

Table IX: Portuguese compounds main lexical-conceptual areas (Ribeiro 2010)

It is also interesting to note that [NA]_N and [[N[PrepN]]]_N patterns are commonly used to denote referents associated to modern and specialized domains of human activity (economy, energy, finance, etc). In these cases, compounds are, mainly, denominations of international circulation, such as *campo magnético* (field magnetic) ‘magnetic field’ or *cadeia de fornecimento* (chain of supply) ‘supply chain’.

5. Headedness

Following Scalise, Fábregas and Forza (2009), the identification of a compound’s head should be based on three different dimensions: categorial, morphological and semantic.

Categorially, the head is the constituent that imposes its categorial features to the whole compound: in (18), the whole’s categorial label is determined by the label of the constituent which functions as the compound’s categorial head.

⁷ In this type we include agents and figurative predications (*pé-de-chumbo* (foot of lead) ‘slow coach’) only applied to human.

- 18) [*vinho*_N *branco*_A]_N [*grande*_A *área*_N]_N
 (wine white) (big area)
 ‘with the wine’ ‘penalty area’

Morphologically, the head of a compound is the constituent that is able to transmit to the compound its morphological features, such as number or gender.

When the categorial and morphological heads coincide (19), they determine the lexical category and the inflectional properties of the outcome:

- 19)
 a) *uma* [*visita*_[-Masc] *-relâmpago* _[+Masc]]_[-Masc]
 (one visit lightning)
 ‘one flying visit’
 b) *duas* [*visitas*_[-Masc] *-relâmpago* _[+Masc]]_[-Masc]
 (two visits lightning)
 ‘two flying visits’

Semantically, the head (a N) functions as a hyperonym of the compound: *jogo de vídeo* (game of video) ‘video game’ and *jogo de salão* (game of salon) ‘parlour game’ denote two hyponyms of *jogo* ‘game’. *Lambe-botas* (lick boots) ‘flatterer’ and *abre-latas* (open cans) ‘can opener’ do not present a semantic head, because none of their constituents function as a hyperonym of the whole and the semantic type of the compound cannot be derived from the semantic properties of its constituents.

The semantic and the categorial heads may or may not coincide: [*poeta*_N *-pintor*_N]_N (poet painter) ‘poet-painter’ has two categorial heads and two semantic heads, [*longa*_A *-metragem*_N]_N (long film) ‘long film’ has one categorial head and one semantic head. Differently, the morphological properties of *um(a)* [*unhas-de-fome*]_N (nails of hunger) ‘skinflint’ are not predictable from those of their constituents. Its denotation is also not predictable from the meaning of the input units: the compound is not a hyponym of none of the constituents.

5.1. Head typology: classes of Portuguese compounds

It is possible to distribute Portuguese compounds into three main classes according to the (in)existence and the position of a head constituent⁸: no head compounds, one head compounds and two-head compounds.

Table X schematizes this classification and presents examples of each type.

No head compounds		[VN] _N : <i>guarda-roupa</i> (keep clothes) ‘wardrobe’
		[V(conj)V] _N : <i>pára-arranca</i> (stop start) ‘stop-start’
One head compounds	Left	[NN] _N : <i>turma-piloto</i> (class pilot) ‘pilot class’
	head	[NA] _N : <i>trabalho infantil</i> (work infantile) ‘child work’
		[[N[PrepN]]] _N : <i>motor de busca</i> (engine of search) ‘search

⁸ Villalva (2000) proposes a summary analysis on the head of Portuguese compounds only based on syntactic criteria. Recent works applied to other Romance languages (Scalise and Guevara, 2006; Scalise, Fabregas and Forza, 2009) stress the need for a wider study, including morphological and semantic criteria.

		engine' [[N[PrepV]]] _N : <i>tábua de engomar</i> (board of starch) 'ironing board'
	Right head	[AN] _N : <i>pronto pagamento</i> (prompt payment) 'cash' Neoclassical compounds: <i>bibliófilo</i> 'bibliophile'
Double head compounds		[AA] _A : <i>afro-americano</i> 'afro-American' [NN] _N : <i>café-restaurante</i> 'café-restaurant'

Table X: Head typology: classes of Portuguese compounds

5.2. Endocentric compounds

Endocentric compounds have a semantic, categorial and morphological head which is modified, specified or complemented by the non head. The compound denotes an entity that is a hyponym of the one denoted by the head. Portuguese endocentric compounds are typically left-headed. They may have the structures displayed on Table XI.

[NN] _N	<i>palavra-chave</i> (word key) 'keyword'
[NA] _N	<i>energia nuclear</i> (energy nuclear) 'nuclear energy'
[[N[PrepN]]] _N	<i>bilhete de identidade</i> (card of identity) 'identity card'
[[N[PrepV]]] _N	<i>saco de dormir</i> (bag of sleep) 'sleeping bag'

Table XI: Endocentric compounds

Though less represented, there are also endocentric compounds with the head on the right, such as *curta-metragem* (short film) 'short film', *malcriado* (badly behaved) 'badly behaved', *hidrofobia* 'hydrophobia'.

There are many compounds, however, that are categorially and morphologically endocentric but semantically exocentric, as those in (20). These compounds present a categorial and morphological head, but this constituent ((20): *pés*) does not function as a semantic head.

- 20) [*pés*_{N [+ Masc, + pl]} *de* *galinha*_{N [- Masc, - pl]}]_{N [+ Masc, + pl]}
(feet of chicken)
'crow's feet'

5.3. Exocentric compounds

Traditionally, exocentric compounds denote a class which is not a hyponym of neither of the classes denoted by the constituents, as in (21).

- 21)
a. [[*caixa*]_{N [- Masc, - pl]} *-de-* [*óculos*]_{N [- Masc, - pl]}]_{N [+ Masc, - pl]} (box of glasses) denotes someone (masc/fem) whose face seems suitable for wearing glasses, and not a

box for glasses.

- b. $[[pele]_{N[-Masc;-pl]} [vermelha]_{A[-Masc;-pl]}]_{N[\pm Masc;-pl]}$ (skin red) ‘redskin’: does not refer to a type of skin, but is an old denomination for native American Indians.

When analyzing Portuguese compounds according to the three dimensions of endo-/exocentricity defined by Scalise, Fabregas and Forza (2009), it is possible to define some groups of compounds which exhibit mixed characteristics.

Some Portuguese compounds are examples of “Absolute Categorical Exocentricity” (Scalise, Fábregas and Forza 2009), because the compound’s characteristics in relation to morphological features, category and semantics are totally different from those of the input units. These cases of extreme exocentricity occur with Portuguese $[VV]_N$ (*vaivém* (go-come) ‘shuttle’), $[VPron]_N$ (*sabe-tudo* (know all) ‘know-all’) and $[VAdv]_N$ (*bota-abaixo* (throw down) ‘intentional destructive criticism’) templates.

There are also cases of compounds which are morphologically and semantic exocentric, but categorially endocentric, as, for instance, $[pele-vermelha]_N$ (skin red) ‘redskin’, $[unhas-de-fome]_N$ (nails of hunger) ‘skinflint’. $[VN]_N$ compounds are also morphologically and semantically exocentric as neither the constituent on the left, governing the internal predicator-argument relation, nor the nominal constituent, on the right, impose their morphological or semantic features to the whole construction⁹.

Table XII demonstrates that categorial, morphological and semantic endocentricity and exocentricity may occur in constructions with different internal structures.

	Morphologically and categorially endocentric	Morphologically and/or ¹⁰ categorially exocentric
Semantically endocentric	$[NA]_N$: <i>guerra civil</i> (war civil) ‘civil war’	
	$[AN]_N$: <i>pequena área</i> (small area) ‘goal area’	
	$[NN]_N$: <i>escola modelo</i> (school model) ‘model school’	
	$[[N[PrepN]]]_N$: <i>cadeira de balanço</i> (chair of swing) ‘rocking chair’	
Semantically exocentric	$[NA]_N$: <i>sangue-frio</i> (blood cold) ‘cold blood’	$[NA]_N$: (um) <i>cabeça-rapada</i> (head shaven) ‘(a) skinhead’
	$[NN]_N$: <i>banho-maria</i> ‘bain-marie’ <i>cara-metade</i> (face half) ‘better half’	$[N_{proper}N_{proper}]_N$: (a) <i>pêra</i> ‘the pear’ <i>maria-antônia</i> (Ribeiro 2010: 24)
	$[[N[PrepN]]]_N$: (o) _[+masc] $[cabeça]_{[-masc]}$ <i>de casal</i> (head of couple) ‘(the) head of the family’	$[[N[PrepN]]]_N$: (um) <i>unhas-de-fome</i> (nails of hunger) ‘skinflint’
	$[AN]_N$:	$[AN]_A$: (um) _[+masc] <i>colega</i>

⁹ The “principle of activation of lexical features” states that a feature only percolates when it has not been satisfied inside the word-internal structure (Scalise, Fábregas and Forza (2009: 24) and sustains exocentricity of Romance VN compounds.

¹⁰ With this formulation we intend to cover (i) morphologically exocentric compounds, such as $[cabeça]_{[-Masc]}-rapada_{[-Masc][+Masc]}$, which are categorially endocentric (*cabeça-rapada* exhibit the same category [N] as its categorial head (*cabeça*), and (ii) other compounds that are morphologically and categorially exocentric, as, for instance, those formed according to $[VPron]_N$ and $[VAdv]_N$ patterns.

	<i>má língua</i> (bad tongue) ‘backbiting’	[<i>boa onda</i> _[-masc]] _[+masc] (good wave) ‘a <u>cool</u> colleague’
		[VV] _N : <i>treme-treme</i> (tremble tremble) ‘jelly’ [VPron] _N : <i>faz-tudo</i> (do everything) ‘handy man’ [VAdv] _N : <i>fala-barato</i> (speak cheap) ‘gasbag’

Table XII: Morphological, categorial and semantic endo- and exocentricity in Portuguese compounds

6. Argument structure and semantic features

In Portuguese compounding the relation between a predicator and its argument is present in [VN]_N structures, where N functions as an internal argument of V (cf. Section 6.1.), in neoclassical structures whose right constituent has a verbal origin (cf. Section 6.2), and in [NA]_N and [N[PprepN]]_N sequences (cf. Section 6.3) where the N on the left is deverbal and has argumental capacity.

In [VN]_N compounds (*conta-gotas* (count drops) ‘dropper’; *amola-tesouras* (sharpen scissors) ‘scissor sharpener’), the two constituents are bound by a predicate-argument relation: the right-hand constituent is the internal complement of the left-hand lexeme, typically a transitive verb.

Neoclassical compounds whose head constituent has a verbal origin, such as *-cida* (*germicida* ‘germicide’), *-ducto* (*oleoducto* ‘oil pipe’), are based on an argumental relation between their constituents. In these compounds the nominal element functioning as the internal argument (the Theme) of the verbal form is on the left.

There are also argumental relations in [NA]_N (*planeamento urbano* (planning urban) ‘town planning’) and [[N[PrepN]]]_N (*acelerador de partículas* (accelerator of particles) ‘particle accelerator’) structures in which the N on the left is deverbal, with predicative capacity to require an internal argument. The constituent on the right, whether an adjective (*urbano*) or a Prepositional Phrase (*de partículas*), fills the propositional space reserved for the Theme of the deverbal constituent.

There is a structural relationship between the subordinate or attributive nature of the compound and its denotational properties: the non-head of subordinate compounds usually denotes a participant affected by the situation described by the predicator, the V on the left (*abre-latas* (open bottles) ‘bottle opener’, *quebra-nozes* (crack nuts) ‘nutcracker’); contrarily, the non-heads of attributive compounds function as a syntactic adjunct, transmitting information such as location, beneficiary, meronymic entities, paths, materials.

Portuguese is less restrictive than English (Lieber 2009: 99) on what concerns [VN]_N semantics: completive compounds denote not only instruments and agents, but also properties, ways of being or behaving or events, filling thematic roles diverse than those typical of external arguments.

The meaning of many compounds (subordinative, attributive and even coordinative) is idiomatic and opaque regarding their compositional semantics. Nevertheless, the encyclopaedic and pragmatic information associated to each constituent underlies the

specific meaning carried by the whole. Coindexation acts in order to conciliate, in a maximal way (Rodrigues and Rio-Torto, forthcoming), the meanings of the constituents with the conventional denotation of the compound¹¹.

7. Compounding and inflection

There is an indelible relation between inflection and head position. Nevertheless, this is not the only criterion used to explain compound inflection. Other criteria must be invoked, such as stress structure, morphological nature of the constituents or grammatical relations between them.

In terms of inflection, Portuguese compounds may be divided into four classes, according to the presence of inflection at the right and/or at the left lexeme, or its external occurrence. In Portuguese, plural is primarily codified by the adjunction of -s to a nominal/adjectival form or to a determiner.

7.1. [$X_{inflection} Y_{inflection}$]: double inflection

Both constituents are inflected, due to mandatory agreement between noun and adjective ((i)-(ii)) or between noun and numeral (iii), or due to the double nominal or adjectival head ((iv)-(v)).

22)

- (i) [NA]_N: *escola(s) primária(s)* (school primary) ‘primary school(s)’
- (ii) [AN]_N: *grande(s) superfície(s)* (big surfaces) ‘shopping(s) centre(s)’
- (iii) [NumN]_N: *primeira(s)-dama(s)* (first lady) ‘first lady(ies)’
- (iv) [AA]_A: *surdo(s)-mudo(s)* (deaf dumb) ‘deaf(s) and dumb(s)’
- (v) [NN]_N coordinate compounds: *médico(s)-escritor(es)* (doctor writer) ‘doctor(s)-writer(s)’

7.2. [$X_{inflection} Y$]: head inflection

Only the head of the compound is pluralized.

23)

- (i) [NN]_N in attributive/modifying compounds: *decreto(s)-lei* ‘decree(s)-law
- (ii) [[N[PrepN]]]_N: *auxiliar(es) de investigação* (assistant of research) ‘research assistant(s)’
- (iii) [[N[PrepV]]]_N: *história(s) de encantar* (story of enchant) ‘fairy tale(s)’

7.3. [$Det_{inflection}[XY]_N$]: exocentric compounds

Exocentric compounds in whose internal structure figures a V do not allow internal

¹¹ Following Rodrigues and Rio-Torto (forthcoming), we defend that the compounds meaning is not always and not only compositional (*lobo do mar* (wolf of sea) ‘experient sailor’). When a literal interpretation is not possible, other semantic devices (metaphorical, metonymic, coercion) are managed, in order to conciliate the meanings of the constituents and the specific meaning of the whole: the speaker conveys the maximal semantic frame of the lexemes, the features associated with the profiles and/or the functions of denotata, as well as the pragmatic purposes words can serve, in order to assure an adequate interpretation. Libben’s (2006:6) perspective confirms that the interpretation of compounds is based not only on storage processes but also on computational ones.

inflection. It is the determiner that expresses the plural. The scope of pluralization is the entire compound.

24)

- i) [VN]_N: *o(s) beija-mão* (kiss-hand) ‘hand-kissing’, *o(s) lava-louça* (wash tablewear) ‘dishwasher’
- ii) [VAdv]_{N/A}: *o(s) fala-barato* (speak cheap) ‘gasbag(s)’
- iii) [VPron]_N: *o(s) sabe-tudo* ‘know-all’
- iv) [V(conj)V]_N: *o(s) sobe-e-desce* (up-and-down)

7.4. [XY]_{inflection}: inflection on the right constituent with scope on the whole

When the compound is a single phonological word, inflection is obligatorily at the end of it. Inflection scopes over the whole structure as in the following agglutinated compounds: *aguardente(s)* ‘brandy(ies)’, *artimanha(s)* ‘trick(s)’.

This group of compounds corresponding to single phonological words also includes those formed by neoclassical roots, such as *anglófilo(s)* ‘Anglophile(s)’, *carnívoro(s)* ‘carnivorous’.

In compounds such as *lusó-sueco(s)* ‘Luso-swedish’, *maníaco-depressiv-o(s)* ‘manic-depressive’, composed by two forms linked by a vowel *[[lus]o[suec]]*, *[[maniac]o[depressiv]]*, inflection occurs at the end of the compound, having scope over the whole. The absence of internal inflection proves that they are morphological compounds, combining roots, not words.

Some [VN]_N exocentric compounds allow a plural mark also on the right constituent (N): *os porta-estandarte(s)* (carry flag(s)) ‘standard-bearer(s)’, *os vira-casaca(s)* (turn coat(s)) ‘turncoat(s)’.

8. Final remarks

Compounds are plurilexematic units whose meaning refers to a delimited reality, with intensional and ontological identity. Each compound unit, whose meaning does not always coincide with the parts’ sum of meanings, is used as a holistic denomination. Despite being plurilexematic, a compound behaves as a single lexical unit, exhibiting different degrees of syntactic and lexical opacity.

Portuguese compounds are characterized by a narrow relationship between internal constituency, headedness and inflectional patterns. Indeed, taking into account the lexical categories of the input units (root, word) and the degree in which the compound’s internal structure departs from syntactic rules, Portuguese compounds are divided into three groups: morphological compounds, morphosyntactic compounds and phrasal compounds. The first class includes roots, the others words.

Morphological compounds are totally resistant to any alteration in its internal structure. Some morphosyntactic and phrasal compounds allow internal inflection.

In any of these classes it is possible to identify compounds based on coordinate, subordinate and attributive internal relations (cf. Table XIII).

	Coordinate	Subordinate	Attributive
Morphological	[<i>agridoce</i>] _A (bitter sweet)	[<i>herbicida</i>] _N (herbicide)	[<i>neurocirurgia</i>] _A (neurosurgery)

	‘bittersweet’	‘herbicide’	‘neurosurgery’
Morphosyntactic	[<i>poeta-pintor</i>] _N (poet painter) ‘poet-painter’ [<i>pára-arranca</i>] _N (stop start) ‘stop/start’	[<i>beija-mão</i>] _N (kiss hand) ‘hand kissing’	[<i>palavra-chave</i>] _N (word key) ‘key-word’
Syntactic	[<i>entra-e-sai</i>] _N (enter and exit) ‘enter and exit’	[<i>abre-latas</i>] _N (open cans) ‘can opener’	[<i>guerra civil</i>] _N (war civil) ‘civil war’

Table XIII: Compounds’ internal constituency: input units and grammatical relations

In relation to headedness, Portuguese compounds are organized into three groups: no head compounds; one head compounds; two head compounds.

All compounds categorially and morphologically exocentric are also semantically exocentric. There are also semantically exocentric compounds which are categorially and/or morphologically endocentric (*lua(s) de mel* (moon of honey) ‘honey-moon’, *unhas de fome* (nails of hunger) ‘skinflint’). The maximum exponent of exocentricity occurs when the compound is morphologically, categorially and semantically exocentric, as [*faz_V-tudo_{PRON}*]_N (do everything) ‘handyman’. Contrarily, the maximum exponent of endocentricity occurs with compounds as *escola modelo* (school model) ‘model school’ or *cadeira de baloiço* (chair of swing) ‘rocking chair’, whose first noun functions as morphological, categorial and semantic head.

Headedness conditions are not only related to internal constituency and internal grammatical relations, but also to inflectional patterns. Exocentric compounds with a V constituent interdict internal inflection. Typically, it is the determiner that expresses the plural, whether a coordinate (*o(s) pára-arranca* ‘stop/start’) or a subordinate one (*o(s) sabe-tudo* ‘know all’).

Morphological compounds – coordinate (*sinojaponês(es)* ‘sinojapanese’), subordinate (*herbicida(s)* ‘herbicide’) or attributive (*hidromassagem(s)* ‘hydromassage’) – require inflection at the end of the compound.

In the other remaining situations, inflection is mainly governed by the head(s): two-inflectional features are required when the compound is double-headed (a scenery that only occurs with coordinative compounds). When the compound is [NA]_N or [AN]_N, inflectional agreement is mandatory between N and A, thus these compounds also exhibit two inflectional features. [[N[PrepN]]_N structures are head inflected.

Table XIV summarizes these tendencies.

Relevant features and inflectional patterns of compounds	Coordination	Subordination	Attribution
(i) exocentricity: [Det _{Inflex} [X Y]]	<i>o(s) pára-arranca</i> ‘stop/start’	<i>o(s) sabe-tudo</i> ‘know all’	<i>o/a(s) fala-barato</i> (speak cheap) ‘gasbag’
(ii) morphological compound [XY] _{Inflex}	<i>sinojaponês(es)</i> ‘Sino Japanese(s)’	<i>herbicida(s)</i> (herbicide) ‘herbicide(s)’	<i>sul-africano(s)</i> ‘South African(s)’

<p>(iii) head(s) inflection [X_{Inflex} Y_{Inflex}]</p>	<p><i>poeta(s)-pintor(es)</i> (poet(s) painter(s)) 'poet(s)-painter(s)' <i>surdo(s)-mudo(s)</i> (deaf(s) dumb(s)) 'deaf(s) and dumb(s)'</p>		<p><i>arma(s) branca(s)</i> (weapon(s) withe(s)) 'sharp weapon(s)' <i>grande(s)</i> <i>superfície(s)</i> (big surface(s)) 'shopping centers'</p>
<p>(iv) head inflection [X_{Inflex} Y]</p>	<p>-</p>		<p><i>fim(s) de semana</i> (end(s) of week) 'weekend(s)' <i>palavra(s)-chave</i> (word(s) key) 'key word(s)'</p>

Table XIV: internal constituency, grammatical relations and inflectional patterns

In what concerns thematic roles, [VN]_N Portuguese compounds can denote not only entities usually functioning as agents and instruments, but also other entities diverse than those typical of external arguments, such as properties, ways of being or behaving, events and locations. Attributive compounds, whose reading is not confined to the predicator-argument relation typical of [VN]_N structures, can refer to a broad universe of beings, things, materials, events, activities. Coordinate appositional compounds are endocentric and denote a single referent, frequently a human-being (*actor-encenador* 'actor-director'), an event (*almoço-debate* 'lunch-debate') or a locative (*bar-discoteca* 'disco-bar'); coordinate copulative compounds (*norte-sul* 'North-South') are exocentric and denote a relationship between their elements.

A global perspective over this interaction between internal structure, headedness and inflection emphasizes some concluding regularities:

i) all compounds including an autonomous verb form are exocentric. Typically, they are externally inflected on the determiner; a few ([VN]_N) exhibit also plural marks on the right constituent.

ii) compounds that include a neoclassical verb form are inflected at the right border of the word.

iii) though all double-head compounds are based on a coordinate relation, some of them, the morphological ones, exhibit inflection only on the right border (*anglo-saxónico-o/a(s)* 'anglo-saxon'), with scope over the whole compound, and some display inflectional features in both heads (*poeta(s)-pintor(es)*). This asymmetry in inflection highlights the different nature of the compound's constituents: in *anglo-saxónico-o/a(s)* the left head is a bound root, resistant to inflection; differently, in *poeta(s)-pintor(es)* both constituents are autonomous nouns, consequently accepting inflection features.

Globally, Portuguese compounds are mainly used to denote objects (*tubo de ensaio* (tube of test) 'test tube'), animals (*pombo-correio* (pigeon mail) 'carrier pigeon'),

plants (*amor-perfeito* (love perfect) ‘pansy’), human beings (*mulher a dias* (woman on days) ‘cleaning lady’). Many human denominations are based on figurative mechanisms such as metaphor and metonymy (*homem-rã* (man frog) ‘frogman’, *pé-descalço* (foot shoeless) ‘tramp’) and with a pejorative label (*mosca-morta* (fly dead) ‘spiritless person’). Though in a smaller number, some compounds denote events (*beija-mão* ‘hand-kissing’, *guerra civil* ‘civil war’, *jantar-comício* ‘dinner-rally’). Many [NA]_N and [[N[PrepN]]]_N compounds are used as technical terms, some of them with an international circulation (*bolsa de valores* ‘stock exchange’).

With the exception of neoclassical compounds, displaying a similar configuration in many languages, Portuguese compounds frequently reflect cultural specificities. This situation is amplified by the pluricontinental scope of this language: indeed, many compounds created in Angola (*pão-burro* (bread donkey) ‘bread flour, baked in a wood oven; lit. ass bread’), Brazil (*pão-duro* (bread arid) ‘miser’), Mozambique (*chapa-cem* (sheet hundred) ‘public open van, whose ride cost (one hundred Meticaís) is displayed on the license plate) or Portugal (*barriga de freira* ‘traditional Portuguese sweet’) are idiomatic codifications, with no literal equivalents in other languages. Contrarily, technical terms are often transparent (*acelerador de partículas* ‘particle accelerator’). Still, it is important to stress (i) the higher frequency of [[[PrepN]]]_N Portuguese compounds, a residual pattern in English, and (ii) the fact that Portuguese is typically a left headed language.

References

- Baptista, Jorge. 1995. *Estabelecimento e Formalização de Classes de Nomes Compostos*. [Establishment and formalization of classes of compound nouns. Master’s Dissertation in Descriptive Linguistics]. Lisboa, Universidade de Lisboa.
- Benigni, Valentina & Francesca Masini. 2009. Compounds in Russian. *Lingue e Linguaggio* VIII (2). 171–193.
- Bisetto, Antonietta & Sergio Scalise. 1999. Compounding: morphology and/or syntax. In Lunella Mereu (ed.), *Boundaries of morphology and syntax*, 31–48. Amsterdam/ Philadelphia, John Benjamins.
- Bisetto, Antonietta & Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* IV (2). 319–332.
- Brito, Ana Maria (2003). Categorias sintáticas (syntactic categories). In Maria Helena Mira Mateus et al. (eds.), *Gramática da Língua Portuguesa* [Portuguese Language Grammar], 323–432. Lisboa, Caminho.
- Cunha, Celso & Lindley Cintra. 1998. *Nova gramática do português contemporâneo* [New grammar of contemporary portuguese]. Lisboa, Sá da Costa Editors.
- Di Sciullo, Anna Maria & Edwin Williams. 1987. *On the Definition of Word*. The MIT Press.
- Faria, André. 2010. Formação de compostos nominais de base livre do PB [Formation of free base-nominal compounds in brasilian portuguese]. In Maria Lúcia Almeida et al., (ed), *Linguística cognitiva em foco: morfologia e semântica do português* [Cognitive linguistics: morphology and semantics of the Portuguese], 205–218. Rio de Janeiro, Publit Soluções Editoriais.
- Gonçalves, Carlos. Forthcoming. *Composição e derivação: pólos prototípicos de um continuum?* [Composition and derivation: prototypical poles of a continuum?]

- Communication I CIEL (Salvador, 17-20 April 2011).
- Guevara, Emiliano & Sergio Scalise. 2009. Searching for universals in compounding. In Sergio Scalise, Elisabetta Magni & Antonietta Bisetto (eds), *Universals of Language Today*, 101–128. Berlin: Springer.
- Libben, Gary. 2006. Why study compound processing? An overview of the issue. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 1–22. Oxford: Oxford University Press.
- Lieber, Rochelle. 2009. A lexical Semantic Approach to Compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 78–104. Oxford: Oxford University Press.
- Lieber, Rochelle & Sergio Scalise. 2007. The Lexical Integrity Hypothesis in a new theoretical universe. In Geert Booij *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting* 1–24. Bologna: University of Bologna. <http://www.morbo.lingue.unibo-it/> (accessed 24 October 2010).
- Mateus, Maria Helena Mira *et al.* (ed). 2003. *Gramática da Língua Portuguesa* [Portuguese language grammar]. Lisboa, Caminho.
- Ralli, Angela. 2010. Compounding versus derivation. In Sergio Scalise and Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 57-75. Amsterdam/Philadelphia, John Benjamins.
- Ribeiro, Sílvia. 2010. *Compostos nominais em português. As estruturas VN, NN, NprepN e NA* [Nominal compounds in Portuguese: VN, NN, NprepN and NA patterns]. München: LINCOM.
- Rio-Torto, Graça. 2006. O Léxico: semântica e gramática das unidades lexicais [The lexicon: semantics and grammar of lexical units]. In Francisca Athayde (coord.), *Estudos sobre léxico e gramática* [Studies on lexicon and grammar], 11–34. Coimbra: Cadernos do CIEG 23.
- Rio-Torto, Graça & Sílvia Ribeiro. 2009. Compounds in Portuguese. *Lingue e Linguaggio* VIII (2). 271–291.
- Rio-Torto, Graça. In print. Derivação [Derivation]. In Maria Antónia Mota *et al.* (eds.), *Gramática do Português* [Portuguese Grammar]. Lisboa: Fundação Calouste Gulbenkian.
- Rodrigues, Alexandra & Graça Rio-Torto. Forthcoming. Semantic coindexation: evidence from Portuguese derivation and compounding. In Pius Ten Hacken (ed.) *Meaning and lexicalization of word formation*.
- Scalise, Sergio, Antonietta Bisetto. 2009. *La struttura delle parole* [Word structure]. Bologna: Il Mulino.
- Scalise, Sergio, Antonio Fábregas & Francesca Forza. 2009. Exocentricity in Compounding. *Genko Kenkyu* 135. 49–84. http://www3.nacos.com/lcj/modules/documents/LSJpapers/journals/135_scalise.pdf (accessed 15 September 2010).
- Štekauer, Pavol. 2005. Compounding and affixation. Any difference? In Wolfgang Dressler *et al.*, *Morphology and its demarcations*, 151–159. Amsterdam / Philadelphia, John Benjamins.
- Vigário, Marina. 2003. *The prosodic word in European Portuguese*. Berlin & New York: Mouton de Gruyter.
- Vigário, Marina (2010). Prosodic structure between the Prosodic Word and the Phonological Phrase: recursive nodes or an independent domain? *The Linguistic Review* 27(4): 485-530.
- Villalva, Alina. 2000. *Estruturas morfológicas. Unidades e hierarquias nas palavras do português* [Morphological structures. Units and hierarchies in Portuguese

words]. Lisboa, Fundação Calouste Gulbenkian.

Villalva, Alina. 2003. Formação de palavras: composição [Word formation: compounding]. In Maria Helena Mira Mateus et al. (eds.), *Gramática da Língua Portuguesa* [Portuguese Language Grammar], 969–983. Lisboa, Caminho.