

In Lingua e Linguaggio VIII(2), 271-281.

Compounding in Portuguese

GRAÇA RIO-TORTO *, SÍLVIA RIBEIRO **

(*University of Coimbra, **University of Aveiro)

In this article, Portuguese compounds are analysed according to different criteria, such as (i) the morphological, categorial and semantic properties of their basic units, (ii) the grammatical relations linking the constituents of a compound, (iii) the patterns of inflection and (iv) the syntactic atomicity and the lexical opacity of compounds.

The problem of the boundaries of compoundhood, namely those existing between compounds and phrasal nouns, is also addressed, as well as the accuracy of the tests adopted to distinguish compounds (especially phrasal or prepositional compounds) from phrases. We assume that in conjunction with the criteria mentioned above, the referential identity of the entity, object, event or property denoted **by the compound** is a crucial dimension for the conceptual unitary of each compound lexeme.

1. INTRODUCTION

In Portuguese, as well as in other languages, compounding consists of the combination of (at least) two lexemes into a larger lexical unit. The general pattern of compounding in Portuguese can be represented by this schema:

1) [[a]_X [b]_Y]_Z

In this diagram, [a] and [b] represent the words, the stems or the roots which make up the compound; X Y and Z represent the major lexical categories associated to [a], [b] and to the compound. The relationship between [a] and [b] is governed by the semantic and the grammatical features of the lexemes. These constituents can be grammatically linked by coordination, subordination and modification or attribution (section 4). This grammatical relation can be implicit or codified by a preposition. The structure of this article is the following: section 2 is devoted to the

delimitation of compounding. Section 3 describes the basic units (roots, stems, words) in Portuguese compounding. Section 4 examines the types of compounds according to their categorial and grammatical structure. Section 5 focuses on endocentricity and exocentricity. Section 6 deals with compounding and inflection. Section 7 examines compounding and argument structure.

The Portuguese data are taken from (i) contemporary dictionaries and on-line corpora (cf. References), (ii) Ribeiro's dissertation data (2006) and (iii) empirical data collected and described by Rio-Torto and Ribeiro for [MorBoComp](#).

2. DELIMITATION: COMPOUNDING AND DERIVATION. COMPOUNDS AND PHRASES

2.1. *Compounding and derivation*

In Portuguese, compounding consists of the combination of two lexemes: two words, or a word and a root/stem. Derivation is characterised by the addition of an affix, a non-lexematic morpheme, to a lexeme — a root or a stem.

Affixes have a fixed position inside **the** word (**çã*+*renova*, **fazer*+*des*) and they cannot combine each other to form a new word (**re*+*ismo*, **in*+*dor*, **des*+*çã*).

In compounding, even when the constituents are bound forms, the combination is more flexible. Some roots are initial combining forms (*geografia*, *psicologia*), others final combining forms (*pesticida*, *polítologo*, *hidrofobia*), and some others can occur in both positions (*filantropia*/ *cinéfilo*, *pedicure*/*ortopedia*). The more or less flexible position of neoclassical roots in Portuguese compounds is similar to all other romance languages. We stress, however, that the set of Portuguese words with a flexible position inside compounds (*bagagem de mão*/*porta-bagagem*; *disco duro* /*cheque disco*; *escola-piloto*/*navio-escola*) is not **universals**. We believe that this aspect, being determined by cultural factors, is specific of each language.

2.1.1. *Compounds and derived words: idiomaticity and semantic compositionality*

Idiomaticity or semantic opacity is not a distinctive criterion to

distinguish between compounds and derivative words as in both cases there may be greater or lesser semantic transparency. The semantic (non) compositionality of the construction characterises compounds as well as derivative words.

The compounds in (2) as well as the suffixed words in (3) are semantically transparent:

(2) *guitarra acústica* ‘acoustic guitar’, *cavalo árabe* ‘Arabian horse’, *ponto de exclamação* ‘exclamation mark’, *cancro de pele* ‘skin cancer’.

(3) *canceroso*, *industrial*, *globalização*, *terrorismo*

The opposite situation coexists with the aforementioned. The meaning of many compounds (4) is not predictable from the meaning of their constituents.

(4) *caixa negra* ‘black box’, *pontapé* ‘kick’, lit ‘the frontend of foot’, *pé de vento* ‘breeze’, lit ‘foot of wind’

The semantics of many affixed words is also not compositional, as may be seen in: *compôr* ‘compose’, different from *pôr com* ‘put with’; *amador* ‘not professional’, and not *amante* ‘lover’ (both derived from the stem of *amar* ‘to love’).

In sum, neither transparency nor semantic compositionality serve as decisive criteria for delimitating derivative words from compounds. Compounds, like derived words, can be ordered in a continuum according to their degrees of lexicalisation.

2.2. *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 (5) illustrate.

(5) *olhos de linxe* ‘eagle eyes’, lit. ‘eyes of linx’, *bilhete de identidade* ‘identity card’, lit. ‘card of identity’

The inputs of these compounds are phrasal units that behave like single words. These single lemma expressed by multiple lexical forms are called phrasal or syntagmatic compounds.

The [NPrepN] configuration may correspond to the compound [[N [PP]]_N or to a noun phrase of the kind [[N [PP]]_{Np}. The [NPrepN] structure with the [N [PP]]_N value has been widely used in Portuguese since the

Middle Ages, but the delimitation between compounds and phrases remains a challenge in the present.

Different criteria — prosodic, semantic, morphological, morphosyntactic, lexical-syntactic – may be invoked in order to distinguish compounds from phrases. The next sections are devoted to their analysis.

2.2.1. *Stress structure*

Stress position is not useful in the distinction of a compound from a phrase in Portuguese. In Germanic languages stress position distinguishes compounds (*'blackbird*) from phrases (NP: *black'bird*): primary 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 stress position is morphologically conditioned. Some compounds combining two roots (cf. 6) are a single phonological word. Other compounds are two phonological words (7), resulting from the combination of a root and a word (7a.) or two words (7b.).

(6)

single phonological words : **agr**ícola ‘agriculture’, **psic**ólogo ‘psychologist’

(7)

two phonological words:

a. **afro-amer**icano ‘Afro-American’, **hidromassagem** ‘hydro-massage’

b. **turma-piloto** ‘pilot team’, **madressilva** ‘honeysuckle’

In the compound nominal structure [N [PP]]_N each noun is a phonological word and the preposition is proclitic.

2.2.2. *Syntactic and lexical criteria of compoundhood*

Lexical integrity and syntactic atomicity of compounds, in the reviewed version of Lieber and Scalise 2007, are two major features of compoundhood. According to these principles, neither deletion nor lexical modification of constituents are allowed in the interior of compounds.

Lexical integrity and syntactic atomicity of compounds are not controversial with morphological compounds (cf. section 3.1.1.) nor with [VN]_N compounds. The problems emerge mainly with prepositional compounds (cf. section 3.1.1.), because they are on the face similar to noun phrases, and the boundaries between both structures remain a

serious issue to linguistic theories.

In order to delimit the frontiers between compounds and phrases Bisetto and Scalise (1999) and Lieber and Scalise (2007) use five tests: (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. A negative result to these tests means that the structures are not phrases but compounds.

Like the Italian compounds tested by these authors, the Portuguese compounds *guarda-costas* ‘bodyguard’, *hidroavião* ‘hydroplane’, *termoacumulador* ‘storage heater/ boiler’, *linhas aéreas* ‘airlines’, *base de dados* ‘database’ do not consent any of the mentioned tests.

Despite the importance of the first four tests, the internal opacity to any lexical changing/insertion remains the main test of compoundhood. In Portuguese, lexical insertion between the constituents or any other modification is normally blocked, as examples (8) illustrate:

(8)

a) **moinho de vento**: *moinho* **antigo* [ancient]/**restaurado* [renovated] *de vento*, *moinho de** *muito/bom* [much/good] *vento*

b) **couve-flor**: *couve* **tenra flor*, *couve* **desta flor*

c) **vinho branco**: *vinho* **completamente branco*, *vinho* **azedo branco*, *vinho* **branquíssimo*

d) **abre-latas**: *abre-* ‘lit. opener of’ **bem*[good] *latas* ‘tins’

As (8) illustrate, in Portuguese compounds the introduction of new material is blocked. Adjectivation, when possible, doesn’t agree with any of the internal constituents of the compound and its scope doesn’t impact individually on the internal terms.

(9) a. *o abre-latas perigoso* or *o perigoso abre-latas*

b. *abre-latas *perigosas* or *abre-*perigosas latas*

Constructions such as (9a) are grammatical, and the adjective *perigoso* ‘dangerous’, in the masculine form, predicates the whole compound. Note that the adjective *perigoso*, since modifying the compound noun in its entirety, is acceptable either in the post-compound or in the pre-compound position. Contrarily, the constructions (9b) *abre-latas *perigosas* or *abre-*perigosas latas* are ungrammatical. In these constructions the adjective *perigosas*[-Masc] ‘dangerous’ would modify *latas*[-Masc] ‘tin’, as proved by gender agreement with the feminine noun *latas*. The fact that such constructions are unacceptable

demonstrates that the adjective can not refer to only one of the constituents of the compound.

Also the reaction of [N[PP]]_N compounds with respect to adjectivation confirms the internal opacity of the Portuguese compounds. Indeed, [N [PP]]_N internal opacity is supported (i) by the gender of the modifying adjectives and (ii) by their scope. Let's take into account the Portuguese equivalent to English 'identity card': *bilhete de identidade*, with gender labels (10) and modified by the adjective *português* (11).

(10) [+Masc][o[+Masc] [bilhete[+Masc] de identidade[-Masc]]]

(11) [+Masc][[o[bilhete[+Masc] de identidade[-Masc]] português
[+Masc]]]

(12) o bilhete de identidade *portuguesa[-Masc]]]

The masculine form "português" is acceptable (11), according gender agreement between the adjective and noun. The scope of the adjective is the whole compound, and the masculine form is determined by the gender of the head.

The feminine adjectival form **portuguesa* (12) is unacceptable because adjectival modification of part of the compound (in this case *identidade*, a feminine noun, as agreement shows) is impossible.

In Portuguese compounding changes in the conditions of determination are also blocked, as they could completely modify the meaning and the reference, mainly when the structure is [NprepN]_N. The compound [+Masc] [fim[+Masc] de semana[-Masc]] 'weekend' refers to a set of days in the break between weeks, that is, the period of rest that goes from the end of the working week until the end of Sunday. In *fim da semana* the presence of the feminine definite article *a* (*de+a>da*) means that the noun *semana* refers to a particular week, which may be determined (con) textually or otherwise located. So, when definite determination is activated, *fim da semana* refers to the last working days of a specific week. So, when a PP is a component of the compound no internal changing of the determination is allowed. Any change in its determination modifies the meaning of the whole structure and has direct influence on its status.

In sum, compounds' constituents can not be specified or modified (cf. 13) by:

(13)

a) **determiners**: *base de *estes dados* 'base of these data'; *bilhete *da identidade*

‘lit. card of the identity’, *carta *da condução* ‘lit. licence of the driving’)

b) **quantifiers** (*base de *mil dados* ‘lit. base of a thousand data’)

c) **adjectives** predicating only one of the constituents (cf. 2.2.3.1.): *abre-latas* **perigosas* lit. ‘an opener of dangerous tin’.

The (im)possibility of lexical insertion (cf. 8), the degree of (un)determination (13) and the scope of adjectivation (11-12) are assumed as important tests of compoundhood, because of the referential effects they produce and the opacity of the whole they reveal. When analysing [NprepN]_N structures in particular, we believe that if internal changing is allowed, these structures are no longer compounds ([N[PP]]_N) and become noun phrases ([N [PP]]_{NP}).

3. BASIC UNITS (WORDS, STEMS, ROOTS) AND TYPES OF COMPOUNDS BASED ON THEM

Portuguese compounds may be based on words (*águas-furtadas* ‘loft’), stems (*limpa_V-neve_N* ‘snowplough’) and roots (*agr-i-cultor* ‘farmer’). Compounds in Portuguese are usually nouns and adjectives. The basic structure of Portuguese nouns and adjectives that can become members of a compound is the following:

(14) Root + (word marker) = stem

3.1. Types of compounds according the [±autonomy] of their constituents

According the [±autonomy] of their constituents, compounds are traditionally divided in two groups: morphological compounds and morphosyntactic compounds. We assume that prepositional compounds [N[PP]]_N represent an other subtype of Portuguese compounds which cannot be neglected.

3.1.1. Morphological, morphosyntactic and prepositional compounds

The recent (2007) linguistic “official Portuguese terminology” (*Dicionário terminológico*, <http://dt.dgicd.min-edu.pt/>), distinguishes between morphological (15a.) and morphosyntactic (15b.) compounds, depending on the bound or free nature of their constituents.

(15)

a. **morphological compounds:** [*agr*]+*i*+ [*cultura*] = [*agricultura*] ‘agriculture’,
 [*hispan*]+*o*+ [*descendente*] = [*hispano-descendente*], [*psic*]+*o*+ [*pata*]
 ‘psychopath’

b. **morphosyntactic compounds:** [*surdo-mudo*] lit. ‘deaf and dumb person’,
 [*guarda-chuva*] ‘umbrella’ lit. ‘protect from the rain’, [*via láctea*] ‘Milky Way’.

Morphological compounds are the result of the combination of one (16a) or two (16b.) bound roots (many of these roots are neoclassical, joined by a linking vowel).

(16)

a. **One bound form:** *afro-americano* ‘Afro-American’, *austro-húngaro* ‘Austro-Hungarian’, *físico-química* ‘Physics and Chemistry’

b. **two bound forms:** *centrípeto* ‘centripetal’, *democracia* ‘democracy’, *herbicida* ‘herbicide’

The linking vowel is *-i-* when the second constituent is [+Linate], such as *aquífero* ‘lit.: water bearer’, *fungicida* ‘fungicide’, *insecticida* ‘insecticide’ and *-o-* in all other situations: *anglófilo* ‘Anglophile’, *astro-física* ‘astro-physics’, *bibliófilo* ‘bibliophile’. However, these regularities are not absolute, as may be seen in *gasoducto*, *oleoducto* ‘gas and oil pipe respectively’, where *-ducto* is [+Linate]. Just as there is *tomaticultura* ‘cultivation of tomatoes’ (Lee, 1987), there is also *aquacultura* ‘aquaculture’.

Morphosyntactic compounds involve two words and are interpreted as the result of a syntactic structure reanalysis (17):

(17)

a. [VN]_N: *abre-latas* ‘tin opener’, *guarda-jóias* ‘jewel box’, *quebra-cabeças*
 ‘puzzle’

b. [NN]_N: *bar-restaurante* ‘restaurant bar’, *couve-flor* ‘cauliflower’, *decreto-lei*
 ‘decree law’

c. [NA]_N: *era cristã* ‘Christian era’, *obra-prima* ‘masterpiece’, *propriedade*
intelectual ‘intellectual property’

d. [AN]_N: *alta-fidelidade* ‘high fidelity’, *belas-artes* ‘fine arts’, *grande área*
 ‘penalty area’

e. [NumN]_N: *dois-cavalos* ‘two horsepower’, *mil-folhas* ‘puff pastry’, *terceira-idade*
 ‘old age’

- f. [AA]_A: *morto-vivo* ‘living dead’, *surdo-mudo* ‘deaf and dumb’, *maniaco-depressivo* ‘manic depressive’
 g. [NV]_N: *sanguessuga* ‘leech’
 h. [VAdv.]_N: *fala-barato* ‘someone who speaks too much’

Syntagmatic or prepositional compounds are words like those on (18), with a [N[PP]]_N configuration. We assume that these structures must be added to the groups previously mentioned. Some of them are the analytical reproductions of Latin compounds (*ludimagister*, *paterfamilias*), in which case inflection was replaced with prepositions.

(18)

- a. [NprepN]_N: *auxiliar de investigação* ‘research assistant’, *caminho de ferro* ‘railway’, *chave-de-fendas* ‘screwdriver’
 b. [NprepV]_N: *máquina de lavar* ‘washing machine’, *ferro de engomar* ‘flat iron’.

These compounds share the grammatical and argumental relations (cf. sections 4 and 7) with morphological and morphosyntactic compounds, and display the same syntactic and lexical features (cf. section 2).

3.1.2. An alternative overview

Concerning the nature of basic units, compounds may be re-assembled as follows:

I. **pluriverbal compounds constituted by (autonomous) words** (19(a)-(h), sometimes connected by prepositions (19(i)-(j))). This class covers morphosyntactic and prepositional compounds.

II. **compounds in which at least one of the constituents is non-autonomous** (*biomassa* ‘biomass’, *indo-europeu* ‘Indo-European’). This group includes the compounds formed by:

- a. **two ‘learned’ bound roots/stems**: *ludoteca* ‘game’s room’.
 b. **a ‘learned’ root and a Portuguese word** (*citodiagnóstico* ‘cell diagnosis’, *hidromassagem* ‘hydro-massage’)
 c. **a portuguese bound root** (*samb-*, from *samba*, *queim-* from *queima*) + **a ‘learned’ stem** (*dromo*), as in: *sambódromo*, the street where the carnival processions take place including samba dance in Brazil, and in *queimódromo*, the streets where the ‘*Queima das Fitas*’ (festival for the students at Coimbra University) procession takes place.

4. TYPES OF COMPOUNDS: CATEGORIAL AND

INTERNAL GRAMMATICAL RELATIONS

4.1. Patterns of compounding according lexical categories

In this section are described the allowed combinations of lexical categories forming a compound. As the data show, the lexical category of the whole is not always identical to the category of the parts. The lexical category of the output can be coincident with the lexical label of the constituents ([NN]_N, [AA]_A), but in other situations there is no equivalence ([NA]_N, [AN]_N, [NumN]_N). In both cases the lexical category of the compound is determined by the lexical category of its head.

When no head exists — when the compound is exocentric —, the lexical category and the gender of the whole ([VN]_N, [VAdv.]_N) emerge in syntax when a determiner is present ([VN]_N *o abre-latas* ‘tin opener’: something that opens tins; [VAdv.]_N *o fala-barato* ‘someone who speaks too much’).

According to the lexical categories involved, the more representative patterns of compounding in Portuguese are the following (19).

(19)

a. [VN]_N: *abre-latas, guarda-jóias, quebra-cabeças*

b. [NN]_N: *couve-flor, bar-restaurante, decreto-lei*

c. [NA]_N: *era cristã, obra-prima, propriedade intelectual*

d. [AN]_N: *alta-fidelidade, belas-artes, grande área*

e. [NumN]_N: *dois-cavalos, mil-folhas, terceira-idade*

f. [AA]_A: where both terms are autonomous (*surdo-mudo*) or where only one of them is a bound form (*sino-americano, ibero-românico*)

g. [NV]_N: *sanguessuga*

h. [VAdv.]_N: *fala-barato*

i. [NprepN]_N: *caminho de ferro, chave-de-fendas, ferro a vapor*

j. [NprepV]_N: *máquina de lavar, ferro de engomar*

k. [NN]_N and [AA]_A neo-classical compounds: *baciforme, centrípeto, democracia, fonoteca, herbicida, hidrofobia*

The productive patterns in contemporary Portuguese are [VN]_N, [NN]_N, [NA]_N, [AA]_A and [NprepN]_N. The schemes [NV]_N, [NprepV]_N,

[VAdv.]_N and even [AN]_N are not productive and [VV]_N, [VPron]_N [AdvA]_A are residual.

In Portuguese the [NprepN]_N remains a very productive template of compounding. Words such as **bolsa valores*, **fim semana*, **comboio passageiros* are unacceptable, because in Portuguese, contrarily to what has been occurring in Italian, the presence of the preposition is still obligatory in the large majority of cases. However, in contemporary Portuguese some [NN]_N compounds without the preposition have started to appear. They are some few terms with universal diffusion, such as: *efeito estufa* ‘greenhouse effect’, *indústria automóvel* ‘car industry’, *imposto automóvel* ‘car tax’, *mapamundo* ‘world map’, *seguro-saúde* ‘health insurance’, which show that Portuguese is also undergoing a change.

4.2. Internal grammatical relations

Compounds in Portuguese can be divided into three classes, determined by the kind of grammatical linking between the constituents: coordination, subordination and attribution. Coordination is based on a symmetrical relation. Subordination implies a predicator-argument completeness and attribution a relationship of modification.

4.2.1. Coordinate Compounds

The constituents of coordinate compounds, which belong obligatory to the same grammatical type, are linked by a copulative operator, either explicit (*whisky e soda* ‘whisky-and-soda’, *sal e pimenta* ‘salt-and-pepper’) or not (*surdo-mudo*, *pateta-alegre* ‘happy fool’). The relationship between the two constituents, both heads of the construction, is symmetrical and additive.

Two subclasses can be identified inside this group: a) **appositional compounds**, such as *trabalhador-estudante* ‘student worker’, where the whole represents a hyponym of both ‘student and ‘worker; and b) **copulative compounds**, such as *Boznia-Herzgovina*, *Alsace-Lorraine* (regions), *Lisboa-Porto* (train), *terra-ar* (míssil) ‘land to air missile’. In these cases the compound is not a hyponym of either constituent. This distinction has visible reflexes in pluralisation (section 6). The major templates of Portuguese coordinate compounds are:

(20)

- a) [NN]_N: *norte-sul* ‘north-south’, *outono-inverno* ‘autumn-winter’, *trabalhador-estudante* ‘student worker’,
b) [AA]_A: *surdo-mudo* ‘deaf and dumb’, *claro-escuro* ‘transition of light to dark’,
austro-húngaro ‘Austro-Hungarian’
c) [VV]_N: *leva-traz* ‘gossip’, *salva-guarda* ‘lifeguard’

4.2.2. Subordinate completive compounds

In subordinate compounds the relation between the two constituents is a predicator-argument relation, a relation of implicit dependence, namely a subordinative one, as the examples (pt.) *guarda-costas* ‘bodyguard’ and (eng.) *taxi driver* show.

The non-head (*costas* ‘back’, *taxi*), is a complement of the (de)verbal head (*guarda* ‘guard’, *driver*), and it represents a THING, in a Lexical Conceptual Structure framework.

Between the two constituents there is a head-argument relation, as in: *condutor de taxi* ‘taxi driver’, *controlo de armas* ‘arms control’, *insecticida* ‘insecticide’, *hidrofobia* ‘hydrophobia’. In Portuguese the head is mainly on the left (*condutor* *de taxi*, *controlo* *de armas*). Nevertheless, neoclassical compounds present the head on the right (*hidrofobia*, *insecticida*) (we underline the head).

These are the templates of subordinate completive compounds:

(21)

- a. [VN]_N *lava-louça* ‘dishwasher’, *abre-latas* ‘tin-opener’
b. [V Pron]_N *faz-tudo* ‘handyman/clown’
c. [NV]_N *sanguessuga* ‘leech’
d. [NA]_N *planeamento familiar* ‘family planning’
e. [N prep N]_N *acelerador de partículas* ‘particle accelerator’, *processador de texto* ‘word processor’
f. Neoclassical: *calorifero* ‘heater’, *cancerígeno* ‘cancer causing’, *carnívoro* ‘carnivorous’, *ovíparo* ‘oviparous’, *pesticida* ‘pesticide’.

The structures (21a,b,d,e) show a prototypical predicator-argument relation, with the argument on the right; (21c) is an atypical case because the argument is on the left. In the cases 21 a,b,d,e the constituent on the right functions as the obligatory complement — the theme — of the constituent on the left. The predicate can be a verb form (a-c), or a deverbal noun (d) referring the action/process denoted by V. In (e) [N

[PP]]_N structures the head is a deverbal noun, and the prepositional phrase [PP] functions as its theme argument (cf. section 7).

The words (f) are borrowed from Greek and/or Latin and/or have been formed according to a neoclassical pattern. The neoclassical units *-cida* ‘kill’, *-fero* ‘take/guide’, *-geno* ‘origins’, *-metria* ‘measure’, *-paro* ‘generate/create’, *-voro* ‘devour’ are the predicate and the left constituent is its internal argument.

4.2.3. *Attributive compounds or modification in compounding: qualifying and classifying subtypes*

Contrarily to subordinate compounds, in attributive compounds there is no dependence relation between the constituents; the constituent on the right restricts the reference of the constituent on the left, specifying and/or modifying its scope, but it does not function as a selected argument. The parts of the compound are linked by a modiferee-modifier relation.

In attributive compounds the non-head, representing a PROPERTY on a lexical-conceptual perspective, specifies an attribute of the head. Then, "attributive" identifies the predicative function of the non-head, normally carried out by:

- an adjective (*bomba atômica* ‘atomic bomb’),
- a noun, with a predicative function (*palavra-chave* ‘keyword’) and a metaphorical meaning, such as *viagem-relâmpago* ‘flying visit’,
- a PP: *cartão de crédito* lit. ‘card of credit’, *óculos de sol*, lit. ‘glasses for the sun’, *visita de medico* ‘a brief visit (as might be performed by a doctor in a hurry)’,
- a numeral (NumN: *quatrocentos* ‘four hundred’).

Two subclasses must be drawn within the group of modifying or attributive compounds. One in which the non-head has a qualifying and/or evaluating function, attributing non defining properties (shape, colour, quality) to the head. The other includes compounds whose non-head constituents have a classificatory or a taxonomic function, denoting a subclass of the class codified by the head.

A **qualifying-evaluative modification** may be expressed by:

(22)

a. an adjective: *mau* ‘bad’, *doce* ‘sweet’, in *mau feitio*, *arroz-doce*, lit. rice sweet ‘rice pudding’,

b. a noun: *chave* ‘key’, *relâmpago* ‘lightening’ or *prodigy* ‘prodigy’ in *palavra-chave* ‘keyword’, *viagem-relâmpago* (‘lightning trip’), *menino-prodígio* ‘child

prodigy’),

c. a Prepositional Phrase: *de palavra* ‘of word’ or *de médico* ‘of a doctor’ (*homem de palavra*, *visita de médico*), which has an essentially qualifying (Bosque 1993), evaluating, and not a classifying function.

d. an adverb: *bem* ‘good’, *mal* ‘bad’ in *bem-humorado* ‘good-humoured’, *malcontente* ‘malcontent’ or *malcheiroso* ‘smelly’.

The major templates of qualifying-evaluative modification/attribution are:

(23)

a. [AN]_N *mau feitio* ‘bad tempered’, *bom gosto* ‘good taste’

b. [NA]_N *horário nobre* ‘prime time’, *arroz doce* ‘rice pudding’

c. [NN]_N *criança-prodígio* ‘child prodigy’, *palavra-chave* ‘key-word’

d. [N[PP]]_N *homem de palavra* ‘man of his word’, *visita de médico* ‘a very quick visit’

e. [Adv. A]_A *bem-humorado* ‘good humoured’, *malcontente* ‘malcontent’,

We assume that there are also **classifying attributive compounds**. In these structures the non-head constituents, classifying the head nouns, have a taxonomic function. The whole compound denotes a subclass of the class codified by the head.

The non-head constituents have ontological and semantic restrictive functions (not qualifying or evaluative ones), involving various kinds of relations — that of ownership, whole/part, location, instrumentality —, depending on the referential nature of what is denoted (24):

(24)

a) *cadeira de rodas* ‘wheelchair’: the noun *rodas* ‘wheel’ denotes a **part** of the chair;

b) *hidroginástica* ‘hydro-gymnastics’, *hidroavião* ‘seaplane’: *hidr-* ‘hydro’ denotes the **locative**;

c) *cartão jovem*: the noun *jovem* ‘youth’ denotes the **beneficiary**;

d) *cardialgia* ‘cardialgia’, *neuralgia* ‘neuralgia’: *cardi-* ‘cardio’ and *nevr-* ‘neuro’ denote the **path** of the pain;

e) *laparoscopia* ‘laparoscope’: *laparo-* denotes the **instrument** through which the vision (*scopia*) occurs;

f) *telefonía* ‘telephony’: *tele-* denotes the **way/channel** through which the sound is transmitted;

g) *biocombustível* ‘biofuel’, *carro a gasóleo* ‘diesel car’: *bio-* and *gasóleo* ‘diesel’ represent the **raw material** of combustion;

The major templates of classifying modificative/attributive compounds

are:

(25)

- a. [NA]_N *bomba atômica* ‘atomic bomb’, *universidade aberta* ‘open university’
- b. [NN]_N *cartão-jovem* ‘student card’, *hidroginástica* ‘hydro-gymnastics’,
- c. [N [PP]]_N *chapéu-de-chuva* ‘umbrella’, *carro a gásóleo* ‘diesel car’
- d. [NumN]_N *terceiro mundo* ‘Third World’, *primeiro-ministro* ‘prime minister’
- e. [AN]_N *alto-relevo* ‘high relief’, *livre arbítrio* ‘free will’

4.3. Crossing grammatical relations

The analysis carried out shows that some grammatical relations underlying compounds are transversal to different patterns of compounding (cf. Ribeiro and Rio-Torto, [in print](#)).

Patterns [NA]_N and [N[PP]]_N are the most multivalent, as they can serve subordinate and attributive relations. The patterns containing a Verb can only support internal subordinate relations. Compounds with a [NN]_N structure sustain only coordinate and attributive relations.

5. ENDOCENTRICITY AND EXOCENTRICITY

The distinction between endocentric and exocentric compounds is correlated with the notion of head. A head of a compound is the element which selects – in formal and/or semantic terms – the other constituent(s) of the structure: the “selected” element(s). In structures such as *abrelatas*, *reestruturação curricular* ‘course restructuring’ or *processador de texto* ‘word processor’, the element on the left “selects” an element which, both in terms of category as well as semantics, is able to complement it.

The head of a compound determines the lexical category (see section 4) and the inflectional properties (see section 6) of the outcome (26):

(26)

- a) *uma* [visita_[-Masc] -relâmpago_[+Masc]]_[-Masc] ‘a flying visit’
- b) *duas* [visitas_[-Masc] -relâmpago_[+Masc]]_[-Masc] ‘two flying visits’

In semantic terms, the head functions as a kind of [hyperonym](#) of the compound. *Jogo de vídeo* ‘video game’ and *jogo de salão* ‘parlour game’ denote two hyponyms of *jogo* ‘game’. Also *astronauta* ‘astronaut’ and *internauta* ‘internaut’ denote two kinds of *nautas* ‘traveller’.

5.1. Endocentric compounds

Generally, endocentric compounds have a head which is modified, specified or complemented by the non head, so that these final products, which have a meaning tending to be compositional, function as hyponyms of the respective head. In Portuguese, there are many endocentric compounds usually with the head on the left. They may have the structures NN (*bolo-rei* ‘a kind of Christmas cake’, *palavra-chave*, *visita-relâmpago*), NA (*cruz-vermelha* ‘Red Cross’, *energia nuclear* ‘nuclear energy’), NPrepN (*bilhete de identidade*, *cartão de crédito*), N PrepV (*máquina de lavar* ‘washing machine’, *ferro de engomar* ‘flat iron’).

There are also endocentric compounds with the head on the right. Their structures include [AN]_N (*curta-metragem* ‘short film’, *boa-fé* ‘good faith’, *grande-área* ‘penalty area’), [Adv A]_N (*abaixo-assinado* ‘petition’, *bem-humorado*, *bem-disposto*, *malcriado* ‘badly behaved’), [Adv V]_V (*bendizer* ‘say well of’, *bem fazer* ‘do good’, *maldizer* ‘say badly of’), and ‘learned’ compounds having one or two neoclassical constituents (*aplaudímetro* ‘instrument for measuring applause’, *filmoteca*, *hidroavião*)

5.2. Exocentric compounds

Compounds included in the group of **exocentric compounds** are those that do not function as a hyponym of neither of the constituents and whose head does not coincide with any of the compound units. See examples (27).

(27)

- a. [+masc] (o) *caixa-de-óculos* which does not denote a box for glasses, but someone whose face seems suitable for wearing glasses;
- b. [+masc] (o) *pele vermelha* ‘redskin’, does not refer to a type of skin, but is an eighteenth and nineteenth century denomination of native American Indians;
- c. [+masc] (o) *banho-maria* ‘bain-marie’, which does not denote a specific kind of bath but a method of preparing food

Exocentric compounds, frequently influenced by strategies of meaning alteration culturally motivated, denote (28):

(28)

- a. **people** that may have the properties expressed by the compound: AN (*bom garfo* ‘someone who enjoys food’), NA (*pele vermelha* ‘red skin’), VN (*vira-casacas* ‘turncoat’), VPron (*sabe-tudo* ‘know-all’)

- b. **instruments** used to perform certain actions: VN (*abre-latas*, *quebra-nozes* ‘nutcracker’)
- c. **products**, such as typical Portuguese desserts/sweets (*barriga-de-freira*, *toucinho-do-céu*), whose denominations are based on metaphorical semantic processes
- d. **states of things, properties, ways of being or behaving**: NA (*roda-viva* ‘a very agitated life’, *sangue-frio* ‘lit. cold blood’, ‘to be too rational or show coldness of character’), NN (*gato-sapato* ‘lit. cat-shoe’, ‘(the fact of) to treat someone badly, to humiliate’), VAdv (*bota-abaixo* ‘intentional destructive criticism’).

Table I demonstrates that endocentricity and exocentricity may occur in constructions with different internal structures.

Internal structure	Endocentric	Exocentric
[AN] _N	<i>belas-artes</i> ‘fine arts’, <i>grande área</i> ‘penalty area’	<i>bom garfo</i> ‘someone who enjoys food’, <i>pobre-diabo</i> ‘poor devil’
[NA] _N	<i>bomba atômica</i> ‘atomic bomb’ <i>guerra-civil</i> ‘civil war’	<i>cabeça-rapada</i> ‘skinhead’, <i>mãos-rotas</i> ‘a very generous person’ lit. ‘hands with holes’
[NN] _N	<i>couve-flor</i> ‘cauliflower’, <i>bomba-relógio</i> ‘time bomb’	<i>banho-maria</i> ‘bain-marie’, <i>cara-metade</i> ‘better half (husband/wife)’
[NprepN] _N	<i>equipa de futebol</i> ‘football team’, <i>jogo de video</i> ‘video game’	<i>rabo de saias</i> ‘man who likes women a lot’, <i>pé-de-cabra</i> ‘crowbar’

Table I: Endocentric and exocentric compounds

6. COMPOUNDING AND INFLECTION

There is an indelible relation between inflection and head position. But the head is not the unique criterion used to explain compound inflection. Different other criteria must be invoked.

It is not difficult to predict the mechanisms of pluralisation in Portuguese compounds, namely when we correlate the inflection process with their internal structure.

In terms of inflection, Portuguese compounds may be divided into four classes, according to the presence of inflection on the right and/or on the left lexeme, or its external occurrence (29). In Portuguese plural is codified by the adjunction of -s to a nominal/adjectival form or to a determiner. These are the conventions adopted here .

(29)

- a. X_{inflection} Y_{inflection}
- b. X_{inflection} Y
- c. inflection [XY]
- d. [XY]_{inflection}

(29) **a. X_{inflection} Y_{inflection}**: both constituents are inflected, due to agreement between noun and adjective ((i)-(ii) or numeral (iii)), or due to the double nominal or adjectival head ((iv)-(v)). In these cases each constituent is followed by its own inflection:

- (i) [NA]_N: *escola(s) primária(s)* ‘primary school(s)’, *obra(s)-prima(s)*
- (ii) [AN]_N: *grande(s) área(s)* ‘penalty(ies) areas’, *grande(s) superfície(s)* ‘shopping (s) centres’
- (iii) [Num ordinal N]_N: *primeira(s)-dama(s)* ‘first lady/ies’, *segundo(s) ciclo(s)* ‘second cycle(s)’
- (iv) [AA]_A *morto(s)-vivo(s)* ‘living(s) + dead(s)’, *surdo(s)-mudo(s)* ‘deaf(s) + dumb(s)’
- (v) [NN]_N in coordinate compounds: *médico(s)-escritor(es)* ‘doctor(s) + writer(s)’, *trabalhador(es)-estudante(s)* ‘student(s) + worker(s)’.

(29) **b. X_{inflection} Y**: this group includes compounds in which only the head (on the left) is pluralised. The structures [NN]_N belonging to this group display an adjunction relation between their constituents.

- (i) [NN]_N in modifying compounds: *decreto(s)-lei* ‘decree(s)-law’, *homem(s)-rã* lit ‘menfrog’
- (ii) [N [PP]]_N: *auxiliar(es) de investigação* ‘research assistant(s)’, *caminho(s) de ferro* ‘rail way(s)’
- (iii) [N PrepV]_N (*máquina(s) de lavar* ‘washing machine(s)’, *ferro(s) de engomar* ‘irons’).

(29) **c. inflection [XY]**: compounds without internal inflection. In these cases, which correspond to exocentric compounds with a [VN] or [VAdv] structure, it is the determiner that expresses the plural. The scope of pluralisation is the entire compound.

- i) [VN]_N: *o(s), este(s)* ‘the, this/these’ *corta-relva* ‘lawn mower’, *o(s) guarda-chuva* ‘umbrella(s)’
- ii) [VAdv]_{N/A}: *o(s) fala-barato* ‘someone who speaks too much and often with poor quality’

(29) d. [XY]inflection: When the compound is a single phonological word, inflection is obligatorily on the end of it. Inflection scopes over the whole structure. This is illustrated by the following agglutinated compounds: *aguardente(s)*, *artimanha(s)*, *fidalgo(s)* ‘noblemen’.

In this group of compounds corresponding to single phonological words are also included those formed by neoclassical roots, such as *anglófilo(s)* ‘Anglophile’, *carnívoro(s)*, *raticida(s)*, *xenofobia(s)*.

In compounds such as *físico-química(s)*, *lusó-chinês(es)*, *maníaco-depressivo(s)*, composed by a bound form, a linking vowel and a free form ([[físic]o[química]], [[lus]o[chinês]], [[maniac]o[depressivo]]), inflection occurs at the end of the compound, having scope over the whole. From a semantic point of view, these compounds are coordinate, conveying two properties (*lusó-chinês(es)*, *maníaco-depressivo(s)*) or two parts of the referent (*(ciências) físico-química(s)* lit. ‘sciences of physics and chemistry’). We claim that the bound nature of the first constituent avoids any inflectional marker, explaining the fact that pluralisation, by default, occurs in the ending position of the compound. Summarizing: pluralisation is governed by a crossing match of several dimensions, such as: stress, head position, bound or free nature of the constituents, type of grammatical relation between them.

7. ARGUMENT STRUCTURE

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 (7.1.), in neoclassical structures whose right constituent has a verbal origin (7.2), and in $[NA]_N$ and $[N[PP]]_N$ sequences (7.3) where the N on the left is deverbal and has argumental capacity.

7.1. $[VN]_N$ compounds

In the $[VN]_N$ compounds the two constituents are bound by a predicate-argument relation: the right-hand constituent is the internal complement of the left-hand lexeme, most frequently a transitive verb. The compound may denote several types of entities, events or properties.

In the following examples (30) V is a transitive verb and the nominal compounds denote:

(30)

- a. **Agent nouns:** *amola-tesouras* ‘scissor sharpener’, *guarda-livros* ‘bookkeeper’
- b. **Instrument nouns:** *abre-latas*, *corta-unhas* ‘nail cutter’
- c. **Event nouns:** *bate-papo* ‘chitchat’, *corta-mato* ‘cross-country run’
- d. **Location nouns:** *quebra-costas* ‘slippery, sloping street’
- e. **Qualifying property** (property of someone who V): *acusa-cristos* ‘delator’, *desmancha-prazeres* ‘wet blanket’

7.2. Neoclassical compounds

An argumental relation is also present in neoclassical compounds where the head constituent has a verbal origin, such as: *-cida* (*germicida* ‘germicide’), *-ducto* (*oleoducto* ‘oil pipe’), *-fero* (*calorifero* ‘heater’), *-fugo* (*febrifugo* ‘febrifuge’), *-geno* (*cancerígeno* ‘cancer causing’). In these compounds the nominal element which functions as the internal argument (the Theme) of the verbal form is on the left.

7.3. $[NA]_N$ and $[N[PP]]_N$ compounds

There are also argumental relations in $[NA]_N$ and $[N[PP]]_N$ structures in which the N on the left is deverbal, with predicative capacity to require an internal argument:

(31)

- a) $[NA]_N$ *planeamento urbano* ‘town planning’ (lit. planning of the town), *reestruturação curricular* ‘course restructuring’
- b) $[N[PP]]_N$ *processador de texto* ‘word processor’, *acelerador de partículas* ‘particle accelerator’.

The lexeme on the right, whether a denominal adjective or a PP, fills the propositional space reserved for the internal argument — the theme — of the deverbal constituent located on the left.

8. Final remarks

Compounds are plurilexematic units whose meaning refers to a particular and delimited reality. A compound denotes a being, thing, event or property, with intensional and ontological frontiers recognised by the members of the community where it is produced. Indeed, each of these plurilexical units is stored in an individual’s mental lexicon as a whole. It is used as a holistic denomination, whose meaning is not always

coincident with the meanings of the parts. Despite being plurilexical, the compounds (*chave-mestra*, *chave inglesa*, *chaves de fendas*) behave as a single unit (syntactically and lexically opaque), usually resisting to any alteration in its internal structure.

A broad approach of Portuguese compounding must take into account all the dimensions described in this article. Portuguese patterns of compounding include neoclassical and Portuguese patterns, originating morphological, morphosyntactic and prepositional templates (we claim that $[N[PP]]_N$ must be considered a template of compounding, as it behaves like the other combinatorial structures). Argumental relations are transversal to different patterns of compounds, such as $[VN]_N$, $[NA]_N$, $[N[PP]]_N$ and neoclassical compounds. In spite of these neoclassical compounds having their head on the right, Portuguese is a left-headed language. The conditions of inflection and the position of the head are prevailing features for the characterisation of compounds.

The lexical-conceptual areas of the compounds and their relationship with the societies where they are used are also an interesting matter. An important set of compounds consists of units of international circulation, some of them being technical terms from different areas (*acelerador de partículas* ‘particle accelerator’, *energia nuclear* ‘nuclear energy’, *cancro de pele* ‘skin cancer’), that become part of our contemporary global specialised knowledge. However, many compounds, particularly those involving metaphors or metonyms related to a particular culture, are specific of each language. As Ribeiro (2006) demonstrates, in Portuguese there are hundreds of compounds based on a figurative codification of the surrounding reality that denote species of animals, plants, and human features. They emphasise the importance of the interface between culture and language and they are certainly a worthy subject for ethnolinguistics.

References

- Baroni, M., E. Guevara & V. Pirrelli (2006), *Sulla Tipologia dei Composti NN in Italiano: principi categoriali ed evidenza distribuzionale a confronto*, in «Atti del convegno SLI 21-23 sept. 2006» (<http://morbo.lingue.unibo.it/guevara/emiliano-guevara-downloadable-papers.html>).
- Bisetto, A. & S. Scalise (1999), *Compounding: morphology and/or syntax?*, in L. Mereu (ed), *Boundaries of morphology and syntax*, Amsterdam/Philadelphia, J. Benjamins, 31-48.
- Bisetto, A. & S. Scalise (2005), *The classification of compounds*, «Lingue e

- Linguaggio», IV (2), 319-332.
- Booij G. (2009), *Construction Morphology and Compounding*, in R. Lieber and P. Štekauer, *The Oxford Handbook of Compoundig*, Oxford, Oxford University Press, 201-216.
- Bosque, I. (1993), *Sobre las diferencias entre los adjetivos relacionales y los calificativos*, «*Revista Argentina de Humanidades*» 9 (1-2), 10-48.
- Dicionário da Língua Portuguesa Online* (<<http://www.infopedia.pt/pesquisa>>), Porto, Porto Editora.
- Dicionário terminológico* (2007), <http://dt.dgicd.min-edu.pt/>
- Guevara, E, & S. Scalise (2004), *V-Compounding in Dutch and Italian*, «Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset», XI, Madrid, I.U. Ortega y Gasset, Also available <http://morbo.lingue.unibo.it/guevara/emiliano-guevara-downloadable-papers.html>
- Lee, Seung-Hwa (1997), *Sobre os compostos no PB*, «*Delta*» 13, 1. <http://www.scielo.br/scielo.php>
- Lieber, R. & S. Scalise (2007), *The Lexical Integrity Hypothesis in a new theoretical universe*, in G. Booij et al. (eds), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting* (Fréjus, 15-18 September 2005). University of Bologna, Available <http://www.morbo.lingue.unibo-it/>
- Ribeiro, S. (2006), *Compostos Nominais em Português: As Estruturas VN, NN, NprepN e NA*. Dissertação de Mestrado em Linguística Portuguesa, Coimbra, Faculdade de Letras da Universidade de Coimbra.
- Ribeiro, S. & G. Rio-Torto (**in print**), *Denominações compositivas de estrutura VN, NN, NprepN e NA: nexos intralexicais*, in «*Actas do XXV Congresso Internacional de Linguística e Filologia Românicas*», Innsbruck, 3-8 Setembro 2007). **Restantes dados**
- Scalise, Sergio (1984), *Generative morphology*, Dordrecht, Foris.
- Scalise, Sergio (1992), *Compounding in Italian*, «*Rivista di Linguistica*» 4, 1, special issue (The Morphology of Compounding), 175-199.
- Villalva, A. (1992), *Compounding in Portuguese*, «*Rivista di Linguistica*» 4, 1, 201-219.
- On-line data base: [/www.linguateca.pt/](http://www.linguateca.pt/)