

ouvrage édité par la

## Cellule de Recherche en Linguistique

Association à vocation scientifique de droit français  
Fondée à Paris en 2001 (Journal officiel n°25 du 23 juin 2001 sous n° 2247)  
Pour la promotion des recherches, de l'information et de la publication en  
linguistique.

### Comité de rédaction

<i>Gerolimich, Sonia</i>	Université d'Udine (Italie)
<i>Ibrahim, Amr Helmy</i>	Université de Franche-Comté & Université Paris-Sorbonne
<i>Martinot, Claire</i>	Université Paris-Sorbonne

## ***PERSPECTIVES HARRISSIENNES***

Textes réunis et édités par

**Claire Martinot, Christiane Marque-Pucheu et Sonia Gerolimich**

Ouvrage publié avec le concours de l'Université Paris-Sorbonne  
EA STIH, 4509  
Equipe d'accueil Sens Textes Informatique Histoire

© CRL, Paris, 2016 -

ISBN 2-9526027-7-8  
EAN 9782952602778

## SOMMAIRE

<i>Présentation du recueil</i>	6
<b>Anne Daladier</b> <i>L'inconsistance de la théorie des opérateurs de Harris : de la fin des grammaires finitistes à une typologie inductive</i>	13
<b>Amr Helmy Ibrahim</b> <i>Ce que les matrices analytiques définitoires doivent à Zellig Sabbetai Harris</i>	36
<b>Gaston Gross</b> <i>L'expression de la finalité en français</i>	48
<b>Christiane Marque-Pucheu</b> <i>Les opérations de réduction dans les expressions formulaires en français : retrouver la langue dans le discours</i>	72
<b>Jean-Pierre Descles</b> <i>Les mathématiques de la Grammaire d'opérateurs de Zellig Harris</i>	83
<b>Bruce Nevin</b> <i>Freeing linguistics from computationalism</i>	106
<b>Sewoenam Chachu</b> <i>Étude comparative des constructions à verbe support en français et en ewé</i>	123
<b>Claire Martinot</b> <i>L'apport de Z.S. Harris à une nouvelle explication acquisitionnelle de la langue maternelle. Enfants francophones de 6 et 10 ans.</i>	145
<b>Sonia Gerolimich &amp; Cristina Castellani</b> <i>L'application des théories harrisiennes dans l'analyse de l'acquisition de l'italien chez des enfants de 8 et 13 ans.</i>	164
<b>Tomislava Bošnjak Botica, Jelena Kuvac Kraljević &amp; Kristina Vujnović Malivuk</b> <i>Complexité linguistique chez les enfants croatophones de 6 et 10 ans dans le cadre de la théorie de Z.S. Harris</i>	188
<b>Danielle Lecman</b> <i>Pour une extension du postulat « la métalangue est dans la langue » : l'explicitation de la situation comme partie prenante de la distribution (exemple de la préposition en devant des « noms de lieu »)</i>	203
<b>Alberto Bramati</b> <i>Les structures syntaxiques contenant c'est. Analyse distributionnelle et traduction en italien</i>	211

<b>Takuya Nakamura</b> <i>On the possible origin of lexicon-grammar tables: speculations from an unpublished manuscript of Zellig Harris</i>	233
<b>Jorge Baptista &amp; Ilia Markov</b> <i>Morphosyntactic processes involving body-part nouns in Portuguese</i>	255
<b>Mirella Cónema &amp; Sara Vecchiato</b> <i>Les composés avec le mot chapeau dans le lexique commun et dans les lexiques satellites</i>	268
<b>Emmanuel Cartier</b> <i>Distributionnalisme et sémantique : état des lieux en traitement automatique des langues</i>	288
<b>Emilio D'agostino, Alberto Maria Langella, Mario Monteleone &amp; Ilaria Maria Villari</b> <i>From Harris Transformational Linguistics to Information Retrieval</i>	314
<b>Giustino De Bueris &amp; Alberto Maria Langella</b> <i>The Legacy of Zellig Sabbetai Harris Renewed: Graph Theory and Transformations</i>	330

Jorge Baptista<sup>1,2</sup>, Ilia Markov<sup>1,2,3</sup>

<sup>1</sup> Universidade do Algarve, Faro, Portugal

<sup>2</sup> Spoken Language Laboratory (L2F), INESC-ID Lisboa, Lisboa, Portugal

<sup>3</sup> Instituto Politécnico Nacional, Centro de Investigación en Computación (CIC), Mexico City, DF, Mexico  
jbaptis@ualg.pt, markoviya@yahoo.com

### Morphosyntactic processes involving body-part nouns in Portuguese\*

#### Résumé

Dans cet article, nous examinons systématiquement l'ensemble (fermé ?) des noms désignant les parties du corps (cœur, main, tête, *etc.*) et les mots morphologiquement associés, en particulier les adjectifs, par référence au portugais européen. Nous démontrons que le cadre harrissien général peut être utilisé pour cartographier la plupart des structures syntaxiques impliquant les noms désignant les parties du corps et leurs dérivés (*crânio/craniano* 'cranium/crânien'), bien que dans certains cas, en raison de la nature technique de ce vocabulaire, certaines classes d'équivalence ne puissent être établies qu'en invoquant l'origine historiquement déterminée de ces termes (par exemple, *coração/cardiaco* 'cœur/cardiaque'). L'article rend compte de ces situations et présente certains cas particulièrement difficiles.

**Mots-clés :** Morphosyntaxe, nom partie-du-corps, adjectif, dérivation, portugais

#### Abstract

In this paper, we systematically investigate the (closed?) set of body-part nouns (*hand, head, hart, etc.*) and their related, morphologically associated, forms, particularly adjectives, with reference to European Portuguese. We demonstrate that the general Harrissian framework can be used to map most of the syntactic structures involving body-part nouns and their related, derived counterparts (*crânio/craniano* 'cranium/cranial'), though in some cases, due to the technical nature of this vocabulary, some equivalence classes can only be established invoking the historically determined origin of the terms (e.g. *coração/cardiaco* 'heart/cardiac'). The paper maps these situations, and shows some particularly difficult cases.

**Keywords:** Morphosyntax, body-part noun, adjective, derivation, Portuguese

#### 1. Introduction

In this work, we focus on a specific class of nouns, human *body-part nouns* (from now on *Nbp*) in (European) Portuguese. This is a finite class of lexical items, that includes both simple (*nariz* 'nose') and compound (*palma da mão* 'palm [of the hand]') nouns; and also several other morphemic units, entering different word-formation processes, and eruditely derived from Greek (*gaster* 'stomach') or Latin (*pulmo* 'lung') roots. Besides, the lexicon of *Nbp* and related vocabulary is composed not only of commonly used designations (*mão* 'hand') but also many specialized technical-scientific terms from Anatomy (*tarsus* 'tarsus'), Physiology (*hematopoiese* 'hematopoiesis'), or general Medical practice (*apendicetomia* 'appendectomy'), which are used almost exclusively in those domains. Since it may be considered to be finite, the lexicon of *Nbp* (and related vocabulary) can be described in

\* Research for this paper has been partially supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2011-2015 and UID/CEC/50021/2013) and the Erasmus Mundus Triple I program, funded by the European Union (grant ref. TRIPLEI2011141).

extension<sup>1</sup>. Our main goal in this paper is to describe morphosyntactic processes concerning *Nbp* (like *nariz* ‘nose’, *palma da mão* ‘palm [of the hand]’) and the relations they hold throughout different parts-of-speech, mainly adjectives (like *nasal* ‘id.’ from the Latin *nasus* ‘nose’) and other noun classes (like *palmada* ‘slap [with the palm of the hand]’). Thus, this work will target different complex relations that, in the framework of Harris (1991) here adopted, may also have a transformational status, besides their historical (diachronic) nature.

One of the motivations for this work is to automatically extract from texts *part-whole relations* (also known as *meronymy*) between *Nbp* and the meronymically related nouns they are associated with (Markov, 2014; Markov *et al.*, 2014). Our purpose is to endow STRING<sup>2</sup> (Mamede *et al.* 2012), a fully-fledged Natural Language Processing (NLP) system for Portuguese, to be able to automatically extract this type of *Nbp*-related meronymic relations from texts. For this goal, a fine-grained description of the *Nbp* lexicon is required, and the complex relation they establish with other lexical categories be established. Another goal is to determine the lexical extension of the morphosyntactic processes here described, as they involve different word-formation phenomena with very limited domains of application.

### 1.1. Body-part nouns, meronymy and *part-whole* semantic relations’ extraction

*Part-whole* relations are a type of semantic relation that holds between two elements in a sentence or text, one that denotes a *part* and another that denotes the *whole*. In this case, we focus on the type of meronymy involving *human* named entities (e.g. a proper name) and *body-part* nouns (also known as *inalienable meronymy*; Boons, Guillet, Leclère, 1976; Guillet, Leclère, 1981). According to our review of the related work and to a relatively recent review of the literature on semantic relations extraction (Abreu *et al.*, 2013), no other mentions on whole-part relations extraction for Portuguese have been identified.

*Meronymy extraction* is (a part of) a NLP task that aims at representing the semantic relations between words expressed in texts, v.g. anaphora and co-reference resolution, semantic roles, entailment, among others.

Arguably, the identification of *part-whole* relations pose a number of problems to natural language processing due to the highly ambiguous nature of the lexicon, the different syntactic constructions involved and even the long-distance gap that has to be bridged in order to establish the meronymy relation between the (often) non-continuous textual units involved. In Portuguese, at least, determinants and modifiers play an important role in capturing the part-whole relation. For example, in (1):

- (1) *O Pedro<sub>i</sub> mordeu uma mão<sub>j</sub>* ‘Peter bit a hand’<sup>3</sup>

since there is an indefinite determiner on the *Nbp*, there is also no obligatory (meronymic) co-reference between *mão* ‘hand’ and the subject *Pedro* ‘Peter’, even though a preferable reading would impose it. Notice, however, that in (2):

- (2) *O Pedro<sub>i</sub> mordeu a mão<sub>i</sub>* ‘Peter bit the hand’

that coreference seems obligatory, though not explicit. Finally, in (3):

- (3) *O Pedro<sub>i</sub> mordeu a sua mão<sub>i,j</sub>* ‘Peter bit the his/her hand’

<sup>1</sup> Non-exhaustive lists of simple and compound *Nbp*, including many idioms involving *Npb*, occurring in the Linguatca *corpora* have been produced and are now publicly available under Project *Esqueleto* (Freitas, 2014): <http://www.linguatca.pt/acesso/Esqueleto/Esqueleto.html>.

<sup>2</sup> <https://string.l2f.inesc-id.pt>

<sup>3</sup> The coreference indices ‘i’ and ‘j’ link the sentence’s constituents, when they refer to the same entity or not, respectively. A word-by-word translation of the examples is provided to highlight the phenomena, and their acceptability is irrelevant.

even if the possessive pronoun *sua* ‘his/her’ is combined with a definite determiner, the same lack of constraint on the *Nbp* coreference shown in (1) can also be found (notice that, in Portuguese, possessive pronouns do not agree with the antecedent in gender number, but are adjective like forms that agree with the noun they determine.). Therefore, a careful consideration of several, apparently unrelated, syntactic and semantic constraints are in order.

In all these cases, the meronymic relation, if it exists, has to be established between the subject noun phrase and a complement, which are not necessarily contiguous. In the ambiguous sentences (1) and (3), the (meronymic) antecedent of the *Nbp* with the non-coreferent interpretation *j* is not even explicit in the sentence, and has to be found in a previous (eventually far away) moment of the discourse.

Several *Nbp*, especially simple nouns, are ambiguous to their meaning: this is the case of *pele(s)*, which can represent both the *Nbp* ‘hair(s)’ and a contraction ‘by\_the’; or the case of *lobo*, a ‘part of the brain’ or a ‘wolf’.

Another issue with *Nbp* is that these nouns can be used metaphorically: *Nbp* properly used for animals are commonly used for human *Nbp*, e.g. *focinho* ‘snout’ instead of *face* ‘id.’. In other cases, the human *Nbp* is actually metaphorically derived from an animal or vegetal literal meaning, e.g. *tronco* ‘trunk’ for *torso* ‘id.’. However, one would like to have a part-whole relation be extracted even if these animal or vegetable *Nbp* are figuratively (and expressively) used instead of the “proper” human *Nbp*:

- (4) *Tira as patas de cima do livro!* ‘Take your paws from the book’

Furthermore, many idioms involve *Nbp*, but for the overall meaning of these expressions the part-whole relation is often irrelevant (Baptista *et al.*, 2014).

- (5) *Maria partiu o coração do Pedro* ‘Mary broke Peter’s heart’

In this idiom, the overall meaning (‘be upset’) has little or nothing to do with the denotation of *coração* ‘heart’, let alone the potential meanings of *partir* ‘break’. Therefore, even though we may consider a part-whole relation between *Pedro* and *coração*, this has no bearing on the semantic representation of the sentence, which could be approximated as for a single-verb predicate – an *operator*, in the sense of Harris (1991): *partir\_o\_corção* (*Maria, Pedro*). In this case, the semantic relation should not, at least, be considered as an ordinary part-whole relation and should be excluded from the semantic parse.

In some cases, a part-whole relation is only implicit, and *Nbp* are involved even though they are not mentioned directly. These are the cases that involve indirect (or lexical) reference, where one can consider that *Nbp* are “hidden” or underlie under certain *Nbp*-derived words. For example, the names of several medical procedures and diseases, where body-part nouns are involved even though not mentioned directly, are formed with (erudit) morphemes directly taken from classical languages (mostly Greek and Latin), like *gastrite*:

- (6) *O Pedro tem uma gastrite por isso não toma café, que dizem faz mal ao estômago*  
‘Peter has a gastritis so he doesn’t drink coffee, for [they] say it is bad for the stomach’.

The part-whole relation is not obvious; however, the coherence of the text relies on identifying that relation. The only reason why this sentence is coherent is because there is a relation between *gastrite* ‘gastritis’ and *estômago* ‘stomach’. We say that the noun *estômago* is “hidden” under the noun *gastrite*, which is an erudite term, directly derived from the Greek root *gaster* ‘stomach’.

### 1.2. Related work

The existence of classes of adjectives derived from *Nbp* was already mentioned for Spanish (Catena, 2006). In her work, the author produces a semantic classification of adjectives. Thus, she distinguishes adjectives related to **physical appearance** (*toothless, leggy*) and **diseases** (*asthmatic, diabetic*). However, the author does not distinguish other subclasses of adjectives, like adjectives related to

medical specialties (*angiológico* 'angiological', *cardiológico* 'cardiological'). Furthermore, the author is not concerned with the morphologic derivation of these adjectives, but rather with their semantic taxonomy. To the best of our knowledge, these classes were not systematically studied for Portuguese, except from some brief remarks on adjectives made by Carvalho (2007).

In her work, Carvalho (2007) proposes a classification of *human adjectives* for Portuguese, that is, adjectives that select human subjects. The author distinguishes a class of adjectives associated with diseases (*artrítico* 'arthritic') or **symptomatic manifestations** (*agoniado* 'sick/indisposed'). The author was not concerned with morphologic processes involved in the formation of these adjectives (respectively, from *artrite* 'arthritis' and *agonia* 'agony'), nor was she addressing the issue of the semantic relation between them (if any). Still, most of the adjectives of this class are neither morphologically nor syntactically related to *Nbp*.

For a few cases, though, an *appropriate Nbp* (Harris 1976: 113-115) can be associated to the adjective. By *appropriation*, Harris (*op. cit.*) defines a word that has an exceptionally high likelihood of co-occurring with another given word, hence contributing with little or no information at all to the sentence, thus being easily zeroable. For example, the subject of *arrepiado* 'goosebump-ed' can be either a human noun or a *Nbp* like *pele* 'skin' or *cabelo/pelo* 'hair':

- (7a) *O Pedro ficou todo arrepiado* 'Peter has goose bumps'  
 (7b) *(A pele, os pelos) do Pedro ficou (ficaram) arrepiado* 'Peter's skin/hair was all goose\_bump-ed'

In fact, the special type of meronymy, *a.k.a. inalienable possession*, triggered by *Nbp* underlies the syntactic-semantic classification and description of some non-completive verb constructions of French (Boons, Guillet & Leclère, 1976), and has also been adopted for the description of Portuguese verbs (Baptista, 2012). Two main classes were set up, depending on the subject or object position of the *Nbp*. Besides the meronymic appropriate reduction, illustrated in (8a)-(8b), other syntactic transformations (Baptista 1992, Correia and Baptista 2007) are characteristic of this verbal constructions, namely, the *dative* (8a)-(8c) (Leclère 1995) and the *locative NP restructuring* (8a)-(8d) (Guillet & Leclère 1981):

- (8a) *O Pedro arranhou o pescoço do João* 'Peter clawed the neck of John'  
 (8b) *O Pedro arranhou o João* 'Peter clawed John'  
 (8c) *O Pedro arranhou o pescoço ao João* 'Peter clawed the neck to John'  
 (8d) *O Pedro arranhou o João no pescoço* 'Peter clawed John in\_the neck'

Baptista (2005) also presents a small list of 30 predicative nouns constructed with support verb *ser de* 'be of'. These nouns constitute, for the most part, nominalizations of adjectives, and their main feature is to have as their subject a particularly appropriated *Nbp*. Hence, *corpulência* 'corpulence' may be associated with *corpo* 'body' as a whole (9), or *palidez* 'pallor' to *pele* 'skin' (10).

- (9) *(O Pedro, o corpo do Pedro) é de uma certa corpulência*  
 '(Peter, the body of Peter) is of a certain corpulence'  
 (10) *(A Ana, a pele da Ana) é de uma certa palidez*  
 '(Ana, Ana's skin) is of a certain pallor'

These authors follow closely Meydan (1996) proposals on the particular distributional constraints and transformational properties of nominal and adjectival constructions involving *Nbp*. Based on these works, we describe the patterns and morphosyntactic processes that we found in Portuguese.

The remaining of this paper is structured as follows: In the next section we present the main morphosyntactic processes involving *Nbp*, describing the derivation of disease nouns, medical procedure and medical specializations, and discussing some complex or problematic cases. The paper concludes with general remarks on the phenomena here presented.

## 2. Morphosyntactic processes with *Nbp*

By morphosyntactic processes we consider not only the historically constraint word formation processes that yield the current vocabulary of the language, but we aim rather at discovering the transformational (paraphrastic) relations, in the sense they are conceived in the Harrissian framework, namely, as syntactic and semantic equivalence relations between sentences. It is clear that many morphological processes involving *Nbp* result from modern terminological needs and those new terms are produced anew (and recently) by recasting Greek and Latin morphemes or roots. In some cases, the paraphrastic relation can only be considered analogically equivalent to other modern processes, since the base structures do not exist in a synchronic perspective.

### 2.1. Diseases

Many nouns designating diseases are obviously related to *Nbp*, *i.e.* they derive from the part of the body affected by an illness. Specialized suffixes are used to denote different types of affections. We start with a relatively regular process, to illustrate the morphosyntactic phenomena and the complexities they may involve.

- Suffix *-ite* '-itis'

The suffix *-ite* '-itis' expresses the meaning of 'inflammation' (or an 'inflammatory state'), and may be attached to a number of *Nbp*. This is the case of the already mentioned case of *gastrite* 'gastritis' but also of several other nouns: *artrite* 'arthritis', from the Gr. *arthros* 'joint'; *bronquite* 'bronchitis', from the Lat. *bronchus* 'id.'; *dermatite* 'dermatitis', from the Gr. *derma* 'skin'; *faringite* 'pharyngitis', from the Gr. *pharynx*, *pharinx* 'throat, pharynx'; *hepatite* 'hepatitis', from the Gr. *hepar* 'liver'; *otite* 'otitis', from the Gr. *otos* 'ear'; *rinite* 'rhinitis', from the Gr. *rinós* 'nose'; *traqueíte* 'tracheitis', from the Lat. *trachea* 'id.', and so forth. Other recent forms can be derived from extant *Nbp*, without having to resource to classical etymons. For example *gingivite* 'gingivitis' (from *gingiva*, Lat. for 'gum tissue'), which was formed upon the Portuguese form *gingiva* 'gingiva'.

All these disease nouns are *predicative nouns* (M. Gross 1981) selecting the support verb *ter* 'have' (durative aspect) or *estar com* 'be with' (more punctual aspect):

- (11) *O Pedro (tem, está com) (artrite, faringite, gastrite, rinite, ...)*  
 'Pedro (has, is with) (arthritis, pharyngitis, gastritis, rhinitis, ...)'

The derivation of these nouns can be directly linked to the base sentence of *inflamação* 'inflammation', with the same support verbs and having the *Nbp* as a (locative) complement:

- (12) *O Pedro (tem, está com) uma inflamação (nas articulações, na faringe, no estômago, no nariz)*  
 'Pedro (has, is with) an inflammation (in the joints, in the pharynx, in the stomach, in the nose)'

Under the Harrissian framework, the paraphrastic relation holding between (11) and (12) can be seen as a reduction of the predicative noun *inflamação* 'inflammation' to a suffix *-ite* '-itis'. This is a process similar to the transformation called *Fusion* (M. Gross 1981), though here it is not a verb and a complement but a predicative noun and a complement (the predicative noun *inflamação* and the locative complement with an *Nbp*) that are involved. In this process, the predicative noun alternates with the suffix *-ite* 'itis', which attaches itself to the *Nbp*, but the transformation keeps all the information of the base sentence. This explains the high redundancy of the sentence of the *Nbp-ite* 'Nbp-itis' predicative noun (11), when a locative complement *em Nbp* 'in Nbp' is added to it, using the same *Nbp* or any other different *Nbp* than the one the *Nbp-ite* was derived from, as it is shown in (13):

- (13) \**O Pedro (tem, está com) (artrite, faringite, gastrite, rinite, ...)* (nas articulações, na faringe, no estômago, no nariz)  
 'Pedro (has, is with) (arthritis, pharyngitis, gastritis, rhinitis, ...) (in the joints, in the pharynx, in the stomach, in the nose)'

In the first case, with the *Nbp-ite* and a locative complement with the same source *Nbp*, the sentence is unacceptable for its high redundancy: the *Nbp* can not be added for it is already present in a derived form, as a predicative noun *Nbp-ite*:

\*O Pedro tem uma faringite na faringe 'Pedro has a pharyngitis in the pharynx'

In the second case, with the *Nbp-ite* and a locative complement with a *Nbp* different from the source of the derived disease noun:

\*O Pedro tem uma faringite no nariz 'Pedro has a pharyngitis in the nose'

the sentence is also unacceptable since no other *Nbp* can be added for the 'argument' position of the sentence's operator, the noun *inflamação* 'inflammation' in its reduced form as the suffix *-ite* 'itis', because that position has already been fulfilled by an form equivalent to a *Nbp*, as an erudite morpheme, *faringe* 'pharynx'.

This is a strong argument to link *synchronously* – and not just *diachronically* – the modern *Nbp* (e.g., *faringe* 'pharynx', *nariz* 'nose') and the erudite equivalent morpheme (v.g., *faring(ô)*, *rin(o)*), irrespective of its formal and historical relation (or lack of), as well as the corresponding sentences, by way of a *transformational process*, thus also linking the predicative noun *inflamação* 'inflammation' and the suffix *-ite* 'itis'.

On the other hand, the locative complement in (12) hints at this not being a base sentence, particularly because of the constraint meronymic coreference between the human subject and the locative prepositional phrase:

\*O Pedro (tem, está com) uma inflamação (nas articulações, na faringe, no estômago, no nariz) do João  
 'Pedro (has, is with) an inflammation (in the joints, in the pharynx, in the stomach, in the nose) of João'

Hence, one can posit that (12) may itself not be the base form but the result of a *locative NP restructuring* (Guillet and Leclère 1981), and should thus be derived from a sentence like (14):

(14) (As articulações, a faringe, o estômago, o nariz) do Pedro (tem/têm, está/estão com) uma inflamação  
 'Pedro's (joints, pharynx, stomach, nose) (has, is with) an inflammation'

Notice also that all the support verbs (*ter* 'have' and *estar com* 'be with', among others) selected by the 'generic' predicative noun *inflamação* 'inflammation' are kept in the derived *Nbp-ite* construction.

Like M. Gross (1981), we propose these type of paraphrastic relations be described by way of a *Fusion* (or *Fusion-like*) transformational process, which reduces the 'generic' predicative noun *inflamação* 'inflammation' to a suffix, *-itis*, binding it to its specialized locative complement with a *Nbp*, and yielding a new, equivalent sentence.

It should be noted that this particular physiological process can only manifest itself in certain body organs or tissues. However, we consider that this issue falls outside the linguistic domain proper, as it describes the extra-linguistic nature of the world, its objects and the relations among them. However, though the morphological process is limited to this domain-specific constraints, it can be considered a *productive process*. Here, *productive* is being used in the sense that any (knowledgeable enough) speaker would unequivocally interpret a newly coined word derived in such a way. For example, a nonexistent *\*ungulite*, from the (previously known) Latin etymon *ungula* 'nail' is immediately interpreted as meaning 'an inflammation in the nail(s)' (and irrespective of that specific human tissue being able to undergo the inflammatory process).

The productivity of this morphologic process is also attested by the ironic formation of words such as *preguicite* 'lazinessitis' and *teimosite* 'stubbornessitis', derived from the predicative nouns *preguiça*

'laziness' and *teimosia* 'stuborness', where the semantic *negative polarity* value of the suffix *-ite* 'itis', namely the *disease* semantic feature, among other connotations, is associated to these type of predicates.

O teu amor incondicional pela Fiat às vezes leva-te à cegueira, que por vezes degenera em teimosite aguda!<sup>4</sup>

'Your unconditional love for Fiat often blinds you and sometimes degenerates into acute stubbornnessitis'

Dir-se-á: um problema de simples distração, se calhar, de mera preguicite de quem orienta o programa, e que, não é a primeira vez, revela tão pouca sensibilidade neste género de critérios.<sup>5</sup>

'One could say: a problem of simple distraction, maybe, of mere lazinessitis from someone who supervises the program and who, not for the first time, reveals so little sensibility in these type of criteria.'

Naturally, these metaphorical extensions of the morphosyntactic and transformational process involving the suffix *-ite* 'itis', though interesting they may be, may not be easy to enumerate nor to model in a NLP system, thus we will not pursue the topic here. The scheme in Fig. 1 summarizes the transformational and morphosyntactic processes described above.

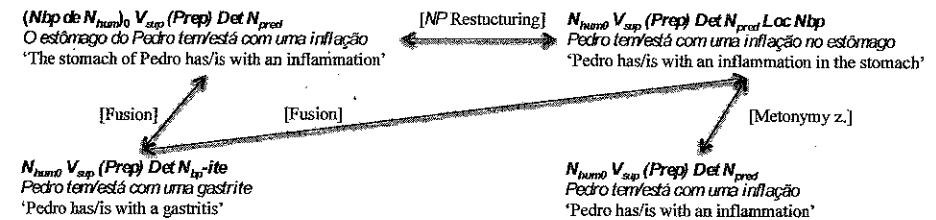


Fig. 1 Morphosyntactic derivation disease nouns from *Nbp* with suffix *-ite*; example: *estômago* 'stomach' and *gastrite* 'gastritis'. *N<sub>hum</sub>* indicates a human noun, *Nbp* is a body-part noun, and *N<sub>pred</sub>* a predicative noun; *V<sub>sup</sub>* is a support verb, *Prep* a preposition, *Loc* a locative preposition, and *Det* a determiner; transformations are represented by arrows and named inside brackets.

Finally, like many other support verb and predicative nouns' constructions, these disease nouns' sentences can be reduced to complex *NPs*, where the predicative noun is the head of the *NP* and its subject becomes a determinative *de N* 'of *N*' complement, as illustrated in (15):

(15) *A artrite do Pedro é grave* (lit: The arthritis of Peter is severe) 'Peter's arthritis is severe'

In this case, the same type of meronymic relation, between *Pedro* and the *Nbp articulações* 'joints' should also be extracted.

Furthermore, like many other nouns, predicative nouns can also yield *de-nominal adjectives*. This is also the case of the *Nbp-ite* nouns, which form *relation adjectives* with suffix *-ico* '-ic', e.g. *gastrítico* 'gastritic' < *gastrite* 'gastritis'. These relation adjectives may, in turn, be used as classifiers on nouns, as in *medicamento antigástrico* 'anti-gastritic medicine', where the prefix *anti-* 'id.' is interpreted as 'against'. This prefix *anti-* can also be added directly to nouns to form adjectives, v.g. *antigastrite* ('anti-gastritis'). Since the predicative noun and the derived adjective may be associated in the lexicon, this could also be used for further representation of the semantic relations involved, v.g. *um medicamento [que é usado] contra a gastrite* 'a medicine [that is used] against gastritis'. This *NP*

<sup>4</sup> <http://forum.autohoje.com/diario-de-bordo/90840-o-mercedes-190e-do-autohoje-vendido-24.html> [07/06/2014]

<sup>5</sup> From the CETEM-Público corpus, extract par=ext1257070-nd-94b-1.

with the noun *medicamento* 'medicine' and prefix *anti-* can be further reduced, by an appropriate zeroing of the head noun, like in (16):

- (16) *O Pedro tomou um (E + medicamento) anti-gastrítico* 'Peter took a (medicine) anti-gastritis'

Finally, let us note the use of these relation adjectives as a human nouns (Carvalho 2007), shown in (17):

- (17) *As azeitonas são particularmente indicadas para gastríticos*<sup>6</sup>  
'Olives are particularly suited for gastritic-s'

This particular nominal use may be derived directly from the base form of the predicative noun *Nbp-ite*, illustrated in (18):

- (18) *Alguém (tem gastrite = é (um) gastrítico)* 'Someone (has gastritis = is (a) gastritic)'

preferably from the sentence with support verb *ter* 'have', due to the durative-permansiive aspectual value associated to the de-adjectival noun. Many other human nouns with the *-ico* suffix, derived from *disease* nouns and not necessarily associated with *Nbp* (at least in a synchronic perspective) are also observed; v.g. *sifilítico* 'syphilitic' < *sífilis* 'syphilis', *diabético* 'diabetic' < *diabético* 'diabetes', *hemofílico* 'hemophilic' < *hemofilia* 'hemophilia', *estrábico* 'strabismic' < *estrabismo* 'strabismus', etc.

Turning now our focus to the NLP modeling problem of representing and processing the semantic relations, namely the extraction of meronymic or *whole-part* relations involved in this processes, it is evident that sentences *ter gastrite = ter uma inflamação no estômago* should yield the same semantic representation. Hence, in the case of the sentence with the predicative disease noun *inflamação* 'inflammation', this requires the adequate identification of the support verb construction, and should yield the semantic dependency:

WHOLE-PART(<subject>,<body-part>)

like, for example:

WHOLE-PART(*Pedro,estômago*)

On the other hand, for the derived predicative noun, such as *gastrite* 'gastritis', etc., the (not always obvious) relation between this noun and the *Nbp estômago* 'stomach' must be made available to the system, so that the same dependency can be extracted. Therefore, extensive lists of disease nouns have been produced, for Portuguese, associating each predicative noun with the *Nbp* it is associated with (and eventually derived from), so that the adequate *whole-part* relation could be extracted. The same type of association is required for adjectives or nominal use of adjectives such as *artrítico* 'artritic'.

#### • *Nbp-ose*

Another example of derived disease nouns is *osteoporose*, 'osteoporosis' (from the Gr. *osteon* 'bone' and *poros* 'pore'), which could be said to mean, literally, 'pores in the bones'. This word form involves the suffix *-ose* '-osis', taken from the Greek and meaning 'illness or morbid state'. The base structure of the *disease* predicative noun is illustrated in the next sentence (19):

- (19) *O Pedro tem osteoporose* 'Peter has osteoporosis'

Though this is an eruditely derived word form, it could also be associated with a base sentence with the noun *poros* 'pores' and a locative complement with the *Nbp* *ossos* 'bones', such as (20):

- (20) *O Pedro tem poros nos ossos* 'Peter has pores in the bones'

We ignore here the fact that *Nbp* is in the plural, and that the singular form exists, but cannot enter a structure like (20):

- \**O Pedro tem um poro no osso* 'Peter has a pore in the bone'

at least not without a particularizing modification on *poro*:

- O Pedro tem um poro infetado no osso* 'Peter has an infected pore in the bone'

This last type of sentences have been analyzed before (Laporte, 1997) as complex sentences, involving an operator-verb (M. Gross, 1981) on a adjectival sentence. The analysis starts from:

- Um osso tem poros # um desses poros está infetado*  
'A bone has pores # one of those pores is infected'

The first sentence is considered to be part of the kernel of language, but from the type of sentences that describe the state of the world, that is, the base information that is involved in the description of sub-languages. It is also responsible for the *part-whole* relation between *poros* and *ossos*. On the other hand, the *part-whole* relation between *Pedro* and *poro* seems to be established on the complex sentence after the reduction of the relative:

- O Pedro tem # um osso tem um poro [que está] infetado*  
'Peter has # a bone has a pore that is infected'

The status of *poros* in this sentence is not clear, as it is a *Nbp* and not a clear-cut predicative noun. In fact, this is not an entirely adequate paraphrase of sentence (19), lacking the specific semantic component that is yielded by the disease noun, which could be located in the suffix *-ose* '-osis'. It could, nevertheless, be taken as the source for the designation (*designatio*), though not of the actual meaning, since in the process of forming this newly coined word some metaphorical process seems to have taken place. This metaphor can be said to be something like 'the bones have too much/large pore like structure as the skin has pores'. A more likely paraphrase would be a sentence like (21):

- (21) *O Pedro tem uma doença nos ossos # Essa doença caracteriza-se por uma excessiva porosidade do ossos*  
'Peter has an illness in the bones' # 'This illness is characterized by excessive porosity of bones'

Notice, on one hand, that *porosidade* 'porosity' is clearly a predicative noun, and that it is derived from the *Nbp* *poros* 'pores'. This predicative noun selects both the support verb *ter* 'have' and *estar com* 'be with':

- Os ossos têm/estão com uma certa porosidade* 'The bones have/are with a certain porosity'

while the verb *estar com* can not appear in (20):

- \**O Pedro está com poros nos ossos* 'Peter is with pores in the bones'

In the same way, sentences like:

- (22a) *Os ossos têm poros* 'The bones have pores'  
(22b) *Os ossos são porosos* 'The bones are porous'  
(22c) *Os ossos têm uma determinada porosidade* 'The bones have a certain porosity'

can be considered an equivalence class, where the *Nbp poros* 'pores' could be defined as a predicative noun. This reintroduces the question of whether to consider sentences like (20) the base form for *osteoporose* 'osteoporosis'. If this is so, then it is not clear which process should be invoked to insert the *-ose* '-osis' suffix into play.

Finally, even if *poros* 'pores' is considered as a plain *Nbp*, that is, an elementary argument in the Harrissian sense, and not an operator (as other, more ordinary, predicative nouns are), there seems to be no word form *\*porose* in the lexicon, which means that the suffix *-ose* '-osis' can only operate once the relation between the morphemes *osteo* (bone) and *poros* (pore) has already been established. Alternatively, the disease noun requires the simultaneous concatenation of the three morphemes, a process that, so far, eludes adequate analysis.

As side remark, notice that the same de-nominal adjective form *osteoporótico* 'osteoporotic' also seems to exist for this morphologic process:

?*um osso osteoporótico* = *esse osso tem/está afetado por osteoporose*  
'an osteoporotic one = this bone has/is affected by osteoporose'

*um doente osteoporótico* 'an osteoporotic patient'

The latter case is similar to *diabetic* 'diabetic', but examples of zeroing of the head noun (?*um osteoporótico* 'an osteoporotic') have not been found.

A similar formation can be found with *osteocondrose* 'osteocondrosis' (from the gr. *osteon* 'bone; cartilage' and *khóndros* 'joint; small, hard and rounded structure'), which could be assigned to a sentence:

(23) *O Pedro tem condros nos ossos* 'Peter has khóndroi in the bones'

even though the noun *\*condro* does not exist in the language (only its diminutive forms *condrulo* 'condrule' e *condrito* 'id.', from Geology, seem to be in use). Notice, however, the use of this Greek source for the *Nbp-ite condrite* 'chondritis', inflammation of a cartilage'), and the derived adjective: *osteocondrítico* 'osteocondritic'.

This *-ose* ending is also used in predicative nouns such *arteriosclerose* 'arteriosclerosis', and *cardioesclerose* 'cardiosclerosis' (from the Gr. *arteria* 'artery' and *kardia* 'heart' and *sclerosis*, that is, a disease in which soft parts inside the body (such as arteries or muscles) become hard. In this case the predicative noun *esclerose* still exists independently, and could be a direct source for the disease noun, though the source sentences are unattested for some body-part nouns:

(24) *O Pedro tem uma cardioesclerose* 'Peter has a cardiosclerosis'  
= ?*O Pedro tem (uma) esclerose no coração* 'Peter has a sclerosis in the heart'

(25) *O Pedro tem uma arterioesclerose* 'Peter has a arteriosclerosis'  
= ?*O Pedro tem (uma) esclerose nas artérias* 'Peter has a sclerosis in the arteries'

We even found a very good example of the later with a more specific body-part noun, the smaller, distal arteries, in site dedicated to linguist problems and curiosities<sup>7</sup>:

cp. *Arterioloesclerose*, como o prefixo indica, tem que ver com processos de esclerose nas artérias mais pequenas, distais – as arteríolas.  
'Arterioloesclerose, like the prefix indicates, has to do with processes of sclerosis in the smaller, distal arteries – the arterioles'

The adjectives *cardiosclerótico* 'cardiosclerótico' or *arteriosclerótico* 'arteriosclerotic' follow the same process described above.

Notice, on the other hand, that the noun *aterosclerose* 'atherosclerosis', ending with the same disease noun, *esclerose* 'sclerosis' is formed of *ateroma* 'atheroma', that is, a lipid formation in the blood vessels. As part of the distributional variation of the subject *esclerose* 'sclerosis', a similar process should then apply in this case, as *ateroma* 'atheroma' can also be analyzed as an *Nbp*.

The observations above could be extended to other suffixed forms, for example *Nbp-ia*, as in *pleurosia* 'pleurisy' and derived adjective *pleurítico* 'pleuritic', an inflammation of the *pleura* 'id.', *miastenia* 'myasthenia' (from gr. *mus* 'muscle' and *astheneia* 'weakness'). With the necessary adaptations, this could also be an adequate analysis for several compound disease nouns, like *câncro de\_ (pele, Nbp...)* 'cancer of *Nbp*', for which the derivation seems more straightforward. Some of these compounds may have an erudite alternative designations (e.g. *melanoma* 'id.', *osteoma* 'id.', *meningioma* 'id.', *linfoma* 'lymphoma', *epitelioma* 'id.'), with an erudite underlying Latin or Greek body-part noun as their base.

## 2.2 Other derivational processes

It would be fastidious to list all the derivational processes that can be found in the language, involving body-part nouns in a more or less direct way. We conclude this section by enumerating several processes (and their related morphemes), which can undergo a similar analysis, based of the paraphrastic relations that can be established between sentences involving *Nbp*.

Several **medical exams** and **chirurgical procedures** are designated by predicative nouns formed on *Nbp* with suffixes such as *-scopia* (from Gr. *scopus* 'see/observe') or *-tomia* (from Gr. *tomo* 'to cut'):

*endoscopia* 'endoscopy', from *endo-* 'inner/inside';  
*lobotomia* 'lobotomy', from *lobo* 'part of the brain';

A similar derivation can be proposed for this type of nouns, involving the predicative nouns *observação/inspeção* 'observation/inspection' and *ablação/excisão* 'ablation/excision, removal':

*O médico fez uma endoscopia ao Pedro* 'The doctor made an endoscopy to Pedro'  
= *O médico fez uma inspeção [ao interior do corpo de] o Pedro*  
'The doctor made an inspection to [the interior of the body of] Pedro'

*O médico fez uma lobotomia ao Pedro* 'The doctor made a lobotomy to Pedro'  
= *O médico fez a ablação de um lobo (cerebral) do Pedro*  
'The doctor made an ablation of a (cerebral) lobe of Pedro'

In Portuguese, the predicative nouns designating **medical exams** and **chirurgical procedures** feature two distinct constructions (Dias de Barros 2014): (i) one, with an **agentive** subject, as in the sentences with *médico* 'medical doctor' as a subject, illustrated above; this construction selects support verbs such as *fazer* 'do', *realizar* 'do', *proceder a* 'do'; (ii) another one, where the subject is the **patient**, but the set of support verbs is different, including *sofrer* 'suffer/undergo':

*O Pedro fez/sofreu uma endoscopia/lobotomia* 'Pedro did/underwent an endoscopy/lobotomy'

We do not address here the eventual (transformational) relations between these two support verb and predicative noun constructions. Since verbs like *fazer* 'do' and *realizar* 'do' (but not *proceder a* 'do') can appear in both constructions, without any additional information it is not possible to determine which is being used in the sentence with *fazer* 'do' above (though a preferential reading may exist).

<sup>7</sup> <https://ciberdividas.iscte-iul.pt/consultorio/perguntas/aterosclerosearteriosclerose/20436>



Notice that, irrespective of this double construction, the same morphosyntactic description, with the necessary adaptations, may apply.

Another interesting class of sentences is the case of the nouns designating **professions and medical specialties**. One group of suffixes that is used to denote professions and specialties is *-ista* 'ist' and *-logia* 'logy', respectively, the latter formed from Gr. *-logia* 'knowledge, science'. These suffixes can be attached to a certain number of *Nbp*. For example, *cardiologista* 'cardiologist'; *cardiologia* 'cardiology' (from Gr. *cardia* 'heart'). As transformations are non-oriented processes, it may be hard to establish which form may constitute a base sentence, but we reason that the name of the specialty could be a good candidate for such derivation:

*O Pedro faz/exerce/pratica cardiologia* 'Peter does/exercises/practice cardiology'  
= *O Pedro é um cardiologista* 'Peter is a cardiologist'

A hint on for the name of the scientific specialty as the source would be a definitional sentence such as *A Cardiologia é a ciência que estuda (as doenças de) o coração* 'Cardiology is the science that studies (the illnesses of) the heart' or *A cardiologia é a especialidade médica que trata (das doenças) do coração* 'Cardiology is the medical specialty that deals with (the illnesses of) the heart. Notice that producing new words by this same process is still very much possible, as the relatively recent word *podologista* (variant: *pedologista*, from Gr. *pous, pod-* 'foot'), a therapist of the feet diseases, demonstrates.

Besides these suffixes, one finds only some few forms ending in *-atra* (*psiquiatra* 'psychiatrist', *fisiatra* and *pediatra*), their compounds, and the corresponding nouns designating the medical specialty. Considering *psiqu-* (from Gr. *psyche*, for 'mind' or 'soul') the base *Nbp*, the origin of the suffix *-atra* (for the profession noun) and *-atria* (for the medical specialty) is not obvious. The *-atra* suffix is considered to be derived from Gr. *iatrós* 'the healer, or the physician'. The *fisi-* morpheme is likely to derive from *physis* 'nature' but the process is much more obscure. Today's interpretation of 'body' or as the biologic processes regulating life, as in *fisiologia* 'physiology' is unclear. The prefix *pedo-*, from the Gr. 'child', is not related with *Nbp* and does not belong to our study set, though we expect the derivational process to be similar.

Interestingly, for the paradigm of *terapeuta* 'therapist', one does not find any professions involving *Nbp*, except perhaps *psicoterapeuta* and *fisioterapeuta*.

### 3. Conclusions

The Harrisian transformational grammar can help to adequately relate the syntactic properties of the *Nbp*-derived predicative nouns like **disease nouns** to the base sentences, where the syntactic-semantic relations between the *disease*, the *body-part* and the *human* possessor, are all overtly expressed. These paraphrastic relations could apply not only to this particular type of predicate, but also be extended to other classes of words, further structuring the lexicon of the language. All these networks of sentences yield the information conveyed in discourse, and confirm the validity of Harris linguistic-theoretical proposals.

These relations are general in nature (e.g., support verb reduction/variation, etc.), but for these classes of nouns, they have limited lexical coverage and are highly constrained for historical/etymological processes. Therefore, they can only be reasonably considered because of their generality – in the sense that they have been observed with many other types of nouns and predicates.

### References

- Abreu, S., Bonamigo, T., Vieira, R., 2013, A review on Relation Extraction with an eye on Portuguese. *Journal of Brazilian Computation Society*, 19(4), 553-571.  
Baptista, Jorge, 1997, Conversão, nomes parte-do-corpo e reestruturação dativa, in Castro, Ivo (ed.), *Actas do XII Encontro da Associação Portuguesa de Linguística* (1996), Lisboa, APL, 51-59.

- 2005, *Sintaxe dos Predicados Nominais com ser de*, Lisboa, Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia, 369 pp.  
Baptista, Jorge, 2012, VIPER: A Lexicon-Grammar of European Portuguese Verbs, in Radimsky, Jan (ed.), *Proceedings of the 31st International Conference on Lexis and Grammar*, Nové Hradý (Czech Republic), September 19-22, 10-16.  
Baptista, Jorge, Correia, Anabela, 2007, Verbal constructions with body-part nouns in Portuguese. *Seminários de Linguística* 5, Faro, Universidade do Algarve, pp. 5-16.  
Baptista, Jorge, Mamede, Nuno, Markov, Ilia, Integrating verbal idioms into an NLP system, 2014, in: Baptista, Jorge et al. (eds.), *Computational Processing of the Portuguese Language, 11th International Conference PROPOR'2014*, São Carlos - SP, Brazil, October 8-10, 2014, Proceedings, *Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence* 8775, Berlin, Springer, 251-256.  
Boons, Jean-Paul, Guillet, Alain, Leclère, Christian, 1976, *La structure des phrases simples en français : Les constructions intransitives*, Paris, Droz.  
Carvalho, Paula, 2007, *Análise e representação de construções adjetivais para processamento automático de texto: adjetivos intransitivos humanos* (PhD Thesis), Lisboa, Universidade Lisboa.  
Freitas, Cláudia, 2014, Esqueleto - Anotação das palavras do corpo humano, Tech. Rep., Linguateca: <http://www.linguateca.pt/aceso/Esqueleto.pdf> (September 2014).  
Gross, Maurice, 1981, Les bases empiriques de la notion de prédicat sémantique, *Langages* 63, Paris, Larousse, 7-52.  
Guillet, Alain and Leclère, Christian, 1981, Restructuration du groupe nominal, *Langages* 63, Paris, Larousse, 99-125.  
Harris, Zellig S., 1964a, Transformations in Linguistic Structure, *Proceedings of the American Philosophical Society* 108, n°5, 418-22 [in Harris, 1970, 472-481].  
1964b, The Elementary Transformations, Excerpts from *Transformations and Discourse Analysis Papers* 54 [in Harris 1970, 482-532].  
1970, *Papers in Structural and Transformational Linguistics*, Dordrecht, D. Reidel Pub. Co.  
1991, *A Theory of Language and Information: A Mathematical Approach*, Oxford, Clarendon Press.  
Laporte, Eric, 1997, Noms appropriés à modifieur obligatoire, *Langages* 126, Paris, Larousse, 79-104.  
Leclère, Christian, 1995, Sur une restructuration dative, *Language Research* 31(1), Seoul, Language Research Institute, Seoul National University, 179-198.  
Mamede, Nuno, Baptista, Jorge, Diniz, Cláudio, Cabarrão, Vera, 2012, STRING: A Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In *Computational Processing of Portuguese, PROPOR 2012*, Demo Session, Paper available at <http://www.propor2012.org/demos/DemoSTRING.pdf>.  
Markov, Ilia, 2014, *Automatic Identification of Whole-Part Relations in Portuguese*. Master's thesis, Faro, Universidad do Algarve/FCBS.  
Markov, Ilia, Mamede, Nuno, Baptista, Jorge, 2015, A Rule-based Meronymy Extraction Module for Portuguese, *Computación y Sistemas* 19(2), 661-683.  
Markov, Ilia, Mamede, Nuno, Baptista, Jorge, 2014, Body-part nouns and whole-part relations, in Baptista, Jorge et al. (eds.), *Computational Processing of Portuguese, PROPOR 2014*, São Carlos, SP, Brazil, October 6-8, 2014, Proceedings, *Lecture Notes in Artificial Intelligence/Lecture Notes in Computer Science*, Vol. 8775, Berlin, Springer, 129-140.  
Meydan, Métié, 1996, Constructions adjetivales, substantifs appropriés et verbes supports. LINX 34-35, Nanterre, Université de Paris 10, 197-210.