

## **Affix Selection and Deadjectival Nouns.**

### **A data-driven approach.**

*Mara Moita (ILTEC), Maarten Janssen (IULA), Margarita Correia (ILTEC/FLUL)*

#### **Abstract**

In this paper, we present a data-driven analysis of the derivational processes in European Portuguese which create property or quality nouns out of adjectives, mostly by means of suffixation. This paper demonstrates which (semantic) restrictions play a role in the applicability of suffixes to base forms. The analysis is based on the SILEX model of derivation, and was done as part of the Semantic Morphological Parser project at the ILTEC institute in Lisbon. The analysis is based on a database of all 3.602 dictionarized deadjectival property noun database, which was created as part of the project. We show that at least for the less productive suffixes for the creation of property nouns, there is a strong preference for each suffix for a specific type of semantics in the adjective the noun is generated from.

Keywords – Word-formation; Derivation; Property Nouns; Lexicology;

## **Affix Selection and Deadjectival Nouns.**

### **A data-driven approach.**

*Mara Moita (ILTEC), Maarten Janssen (IULA), Margarita Correia (ILTEC/FLUL)*

#### 1. Introduction

In many Indo-European languages, and particularly in Romance languages, there are various competing mechanisms for creating (new) words by means of affixation. For instance, there are several highly productive suffixes for creating property or quality nouns from qualitative adjectives in Portuguese. In some cases, it is even possible to use more than one suffix with the same base form, without strong differences in meaning between the different derivatives. However, despite the high productivity of these suffixes, not all of them can be used on all adjectival base forms.

As part of the *Semantic Morphological Parser* project (henceforth SMP), a data-driven analysis is being made of the restrictions that play a role in the formation of deadjectival nouns in European Portuguese, with a focus on semantic restrictions. This analysis is based on the SILEX model by Danielle Corbin (1987; 1991; 1997), a model in which the meaning of the derived words is analysed in terms of the semantic contributions of the root, the suffix, and the affixation process itself.

The final goal of the analysis is to give a full overview of the relative productivity of the various processes for generating deadjectival nouns, based on the 37.176 adjectives in *MorDebe*. *MorDebe* is a general language lexicon annex morphological database for Portuguese, which can be consulted via the Portal da Língua Portuguesa. The main objective of this paper is to show that the semantic characteristics of adjectives play a fundamental role in the construction of deadjectival property nouns (henceforth DPN), a hypothesis already suggested by Correia (1999). This will be demonstrated using a preliminary data-driven analysis of the lexicalized property nouns

in European Portuguese.

## 2. Semantic Morphological Parser

The Semantic Morphological Parser project is an ongoing project, funded by the Portuguese research council (FCT – PTDC/LIN/66037/2006), which has the following main goals:

- (1) the extension of a set of databases of inherent inflection relations (and possibly other derivations) in MorDebe, which will be made available to the public. The database of property nouns (henceforth PN) described here is only one of the databases under development;
- (2) the study of the lexicalized relations stored in these database, in order to find regularities in the type of derivations allowed on each type of base form, and model these regularities (and restrictions) in terms of the morphological theory developed by the SILEX group in Lille;
- (3) the construction of a prototype morphological parser that takes at least some of these restrictions into account.

The morphological databases will be made available via the *Portal da Língua Portuguesa* ([www.portaldalinguaportuguesa.org](http://www.portaldalinguaportuguesa.org)), a web-site dedicated to the distribution of mostly lexical information regarding the Portuguese language (Janssen 2007). On that site, the databases will be made available as dedicated dictionaries, much like the dictionary of deverbal nouns and the dictionary of toponyms and demonyms, which are already available. The addition of the data on PN will increase the information contained in *Mordebe* and the systematic study of historical, morphological and semantic aspects related with different types of derivational processes.

## 2. Property Nouns

## 2.1. Grammatical Properties

A property noun is a noun that is used to name a quality or state. Most of these PN are deadjectival property nouns, where the noun is derived from an adjective, and relates to the same property the adjective itself denotes. A typical example for English language is the noun *braveness* from the (base) adjective *brave*. However, there are also PN that are not derived from adjectives. There are simple PN, such as the noun *wealth*. Most of these “primitive” PN have a derived adjective associated with them, which largely fulfils the same role as the base adjective. In the case of *wealth*, the derived adjective is *wealthy*. Furthermore, there are PN which are derived from nouns and some derived from verbs. In the study presented here, we will focus entirely on the DPN, and particularly on the DPN formed by suffixation, which are the most frequent, the most typical and the most regular property nouns (cf. Correia 1999).

DPN form a rather homogenous group of nouns: they are always abstract nouns, which are in principle uncountable. Furthermore, they all display the characteristic properties listed below.

1. Predicativity: adjectives that have a corresponding DPN can occur in predicative constructions like NP + *to be* + ADJ, e.g. *braveness* – *John was brave*;
2. Multiple occurrence: every entity that has the property of being *pale*, manifest the quality of *pallor*;
3. Ontological dependence: properties can only occur in the presence of an entity – if there is *pallor* then there has to be an entity which is *pale*;
4. Polysemy: any PN can be used to denote the prototypical entity that manifests the quality in a very clear or relevant way: a *beauty* is a person (men or woman) that has a lot of beauty, an *honour* is a fact or action that provides honour in a relevant way.

Although all of these properties apply to all DPN without significant exceptions, these

characteristics are not intended as completely defining the notion of a DPN. A detailed description of the selection criteria used in the current study for considering a noun a DPN will be given in chapter 3.

## 2.2. Affix Restrictions – Morphological and Semantic

There is a rich and diverse set of suffixes which can be used to derive PN from adjectives. But there are several factors restricting which suffix can be used with which adjective, as was observed by Correia (1999). The most remarkable factors are: the etymology of the adjective, its structure, the attitude of the adjective towards the entity it refers to, the linguistic register of the adjective, and the type of quality which the adjective names. The types of qualities that are significant in these restrictions are: objective, stereotypical, sensory, intellectual or emotive qualities. Some examples of these restrictions are given below.

### 1. Cases in which the morphological structure of the adjective restricts the applicable suffixes:

- adjectives ending in *-vel* can only have property nouns in *-idade* and *-ismo*:  
*imóvel* (still) > *imobilidade* / *imobilismo*;
- suffixes *-ez* and *-eza* only select either simple or prefixed adjectives as bases: *\*ocioseza* (oci-os-eza – idle-ness);

### 2. Cases where the application of suffixes can be restricted by etymology:

- suffix *-ia* selects in principle only Greek bases:  
*análogo* (from greek, analogous) > *analogia* (analogy);
- suffix *-idade* selects predominantly Latin and vernacular words:  
*claro* (Latin *clarus*) (clear) > *claridade* (clarity);
- suffix *-ez* selects only Latinisms:  
*estúpido* (Latin *stupidus*) (stupid) > *estupidez* (stupidity).

Apart from morphological and etymological restrictions on the applicability of the suffixes, the choice of the suffix also affects the typical meaning of the DPN. Some examples are given below, taken from Correia (1999).

3. Most property nouns derived with *-ice* present axiological (negative) values:

*asnice* (dumbness), *cobardice* (cowardice);

4. Most property nouns derived with *-ice* and *-ismo* refer to human behavioural characteristics: *malandrice* (the fact of being crook), *beatice* (religiosity);

5. Nouns in *-ia*, *-idade*, *-ismo* and *-ez* are often scientific and technical terms, given the fact of denoting properties in an objective manner:

*imagiologia* (imagiology), *aartesianismo* (artesianism);

6. Suffixes *-idão* and *-ura* focus on physical (or sensory) properties:

*pretidão* (blackness), *espessura* (thickness).

### 3. Methodology

#### 3.1. Creation of the Data-Base

The basis for the analysis presented in this paper was a lexicographically based study of PN in Portuguese. That is to say, the DPN taken into account in this study are those listed and defined as PN in the major general language dictionaries for Portuguese. The dictionaries used for the analysis are on the one hand the Brazilian dictionaries Houaiss and Michaelis, and, on the other hand, the Portuguese dictionaries DPLC (2001) and GDLP (2004). The Brazilian dictionaries were selected because they were available in a machine-readable format; the Portuguese dictionaries for the manual confirmation of the automatically extracted data.

Because of the semantic aspects present in the definition of a DPN, the first criterion for inclusion of a DPN candidate in the database was the semantic definition provided by the dictionary, and not, for instance, their etymological information: when

the dictionary characterizes the noun as etymologically derived from an adjective, but the semantic definition does not fit the standards for a PN, the candidate was not included in the database.

The initial hypothesis was that each dictionary uses a restricted set of fixed formulae to define words as quality nouns, such as “*qualidade de X*”. To find out which were these fixed formulae used in both dictionaries under analysis, a selection of known PN was manually verified. A set of typical definitional patterns was established, to automatically perform a Lexicosyntactic Pattern Extraction (LSPE) process (Hearst 98), thus extracting potential DPN. Table 1 shows an abstraction of the patterns used in the LSPE process, with an example from the dictionary for each pattern.

#### *Houaiss*

Qualidade de [ser] [ADJ]	acronicidade – qualidade deacrônico
Característica de [ser] [ADJ]	lindeza - característica de lindo
Natureza do que [ser] [ADJ]	bastidão - natureza do que é basto

#### *Michaelis*

Qualidade de [ser] [ADJ]	esperancismo - qualidade de ser
Característica de [ser] [ADJ]	especificidade - característica de ser
O fato de [ser] [ADJ]	inatismo - o fato de ser inato

Table 1 – Examples from definitional patterns

Using these patterns, we then automatically extracted all nominal entries matching them, and stored the list of resulting nouns and their respective adjectives in a database. This search resulted in a list of around 5,000 DPN candidates.

### 3.2. Candidate Verification

Lexicosyntactic pattern extraction is never a perfect process; there are always *false candidates* in the list. False candidates in this case are those entries in the dictionary where the wording in the definition is similar to that of a DPN (matching one of the patterns), but for which the noun does not actually correspond to a DPN. An example is the definition of *magma* below, where the pattern will yield a relation between *magma*

and *pasta*.

magma, *sm* (gr *magma*) 1 mistura de matéria mineral ou orgânica em estado de pasta fina.

Because of the presence of the false candidates, the second step in the process of the construction of a DPN database was to manually verify all the elements on this candidate list, keeping only those nouns with a clear synchronic correspondence between morphological structure and meaning, to indeed be DPN.

After cleaning out the false candidates from the list, the resulting database consisted of just over 3,000 DPN.

### 3.3. Manual Extension

The definitional treatment of DPN in dictionaries is not completely homogeneous, much like the treatment of any other class of words. Although most of definitions for DPN use the patterns like those in Table 1, there are also many cases in which the definition deviates from them. Two examples are given below: the definition of Houaiss for the word *cobardia* (cowardness) in (3) uses a periphrasis (“an action that denotes fear or perversity”); and in the case of *adversidade* (adversity), the definition (1) is done in terms of synonymy, using other DPN, which also mean “adversity”, but which are derived from other adjectives.

*cobardia sf (cobarde+ia)* 1. medo, pusilanimidade. 2. acanhamento, timidez. 3.

Ação que denota medo ou perversidade . 4. ânimo traiçoeiro. Var: *covardia*.

Antôn: *coragem, valentia*.

*adversidade sf (lat adversitate)* 1. desgraça , infelicidade , infortúnio , sorte adversa.

2. contrariedade, contratempo, revés.

Thus, many PN were not captured by the semi-automatic harvesting process. Therefore,



additional words were harvested in the following way: departing from data in *MorDebe* we verified all those nouns with a morphological structure typically related to a DPN (i.e. those structures frequently occurring in the list of words harvested thus far), but which were not harvested by the LSPE process. Those words were then manually verified in two Portuguese contemporary general language dictionaries – DPLC (2001) and GDLP (2004), to verify whether they were indeed DPN. This verification was not purely based on the definition pattern: it also took into account whether the morphological structure and meaning of the word match that of a DPN, independently of the actual formulation used in the dictionary.

Etymological (or historical) criteria were not taken into account. For instance, although *urbanidade* (urbanity) is defined in Houaiss as coming from the Latin word *urbanitate*, it was still kept in the database as a DPN resulting from the suffixation of *urbano* with *-idade*. related to the adjective *urbano* (urban). + *-idade*.

This second harvesting process yielded around 500 additional DPN, which were added to the existing database, producing a final set of 3,500 DPN, which will be under analysis in the remainder of this paper.

### 3.4. Property Nouns Database

In principle, the DPN database contains the complete set of lexicalized DPN in the dictionaries used as reference. However, it is worth referring two restrictions in the selection. Firstly, only DPN with deadjectival morphological structure were considered; this means that we did not consider simple PN, such as *saúde* (health), nor words for which there is no explicit derivational process for the generation of the noun from the adjective: for instance, the adjective *acuidade* (acuity) etymologically relates to the adjective *agudo* (acute), but the structure of this noun cannot be explained on the bases of the form of the adjective, since it was instead adopted from the French word *acuité*.

Secondly, only those DPN for which the adjectival base was attested in the dictionaries were considered: a noun like *diglossia* (“the fact of having two tongues or two languages”) was not considered, because although it could be said to be derived from a possible adjective *diglossa* (“bilingual”), there is no such adjective registered in any dictionary of the Portuguese language.

#### 4. Analysis

##### 4.1. Theoretical framework

For the analysis of the derivational processes of DPN creation, we used the *constructional morphologic model* developed by Corbin (1987; 1991, 1997), and used to describe Portuguese PN by Correia (1999). This framework is a synchronic lexical theory modelling the notion of “lexical grammaticality” by means of an unified structure and interpretation for word construction, which accounts for both attested and possible but unattested words. A primary focus in this framework is to determine the restrictions that play a role in the process of word-formation.

A fundamental hypothesis of the SILEX model is that there is a single hierarchical set of rules and principles to construe words, a subset of which accounts for all the observable irregularities in lexically attested words. The rules of SILEX associate a derivational process (suffixation), a base-word category (here an adjective), and the set of operators (affixes) involved in the process. They establish which processes can be used on which types of base words; the meaning of the derived word can be determined by the conjunction of the meaning of the base-word, the meaning contributed by the affix, and the meaning resulting from the process itself.

With respect to DPN, a SILEX analysis tries to establish for each of the DPN-forming suffixes, which types of adjectives it can be used with. Furthermore, it tries to establish which kind of meaning each DPN-forming suffix can contribute.

#### 4.2. Methodology of the analysis

To analyse the data in the database along the lines of the SILEX model, it is in principle necessary to verify every possible characteristic of the base, of the suffix and of the relation between base, suffix and morphological rule. Since this is a huge time consuming work, we started from the conclusions drawn in Correia (1999), and continued the analysis from there. This means that on the one hand, we verified whether the conclusions presented there are indeed confirmed by the data, and, on the other hand, studied other patterns similar to those found there in order to look for additional restrictions on top of those presented there.

As a first step in the analysis, we did not take all the 3.600 DPN into account, but rather took a sample set of 40 nouns for each of the productive suffixes (as listed in the next section) and tested the restriction for this sample set. These 40 nouns were selected at random and should present a representative set of nouns. This sample set was then used to verify the distribution of the 8 productive suffixes over all the types of base adjectives listed in section 2.2, as well as a number of other adjective classes. That is to say, we verified which suffixes may be used to form nouns of properties (behavioural, physical, etc); which suffixes combine with adjectives with a negative connotation; etc.

As a second step in the analysis, whenever a correlation was found between the suffixes and a type of adjective (such as the fact that the suffix *-ia* only applies to Greek based adjectives for instance), we verified whether that tendency was an accidental property of the sample set, or whether it also apply to the complete set of DPN.

As a final step, for all the real tendencies in the database of DPN, we tested whether this was merely a property of the lexicalized DPN, or whether it would really be impossible to apply to suffix *-ia* to adjectives that are not derived from a latin word.

The analysis is still in progress, the first stage (verification in the sample set) is

almost complete, and the second stage (verification in the entire DPN database) is well under way. In this paper however, we merely present the correlations found in the sample set, and only those conclusions relating to the semantic typology of the adjectives.

#### 4.3. Property Noun Structures

Although we only took deadjectival PN into account in the analysis following in the next few sections, we did keep a database of all the other types of PN as well. The analysis of this database, with 370 PN, shows that there are basically five ways of forming PN, as shown below. The formation processes are:

- (1) simple nouns for which the corresponding adjective is derived from the noun:

*fama*<sub>N</sub> (fame) > *famoso*<sub>ADJ</sub> (famous);

*honra*<sub>N</sub> (honour) > *honrado*<sub>ADJ</sub> (honourable);

- (2) deverbal nouns which also gave rise to a deverbal adjective, generally from the past participle form of the verb:

*adequar*<sub>V</sub> (to adequate) > *adequação*<sub>N</sub> (adequacy) > *adequado*<sub>ADJ</sub> (adequate);

- (3) denominal nouns for status, which can also name the identifiable property conferred by the status:

*anonimato*<sub>N</sub> (status of anonymous people) – “the fact of being anonymous”;

- (4) compound nouns, which are used as scientific terms:

*anisogenia* (*aniso+geno+ia*) ( “property of presenting different inheritances”);

- (5) denominal collective nouns, which can also be used to denote the stereotypical behaviour of its members:

*vadiagem*<sub>N</sub> (group of punks) - “quality of punk (behaviour, way of life)”.

The relative frequency of these types of PN with a non-adjectival base are shown in table 2. The numbers in table 2 are complicated by the fact that there are nouns which

could be counted under more than one class. Although there are hence other types of PN, the deadjectival property nouns are by far the most common, constituting more than 90% of the property nouns.

Simple nouns	20%
Deverbal nouns	26%
Denominal nouns	23%
Compound nouns	23%
Denominal collective nouns	9%

Table 2: Total occurrences of non-adjectival PN

The DPN in the database are without exception all formed by way of suffixation. As said before, there are several productive suffixes in Portuguese for the creation of DPN. The relative productivity of the suffixes used for the creation of property nouns is shown in table 3, based on the 3.602 nouns in the database.

-idade	1890	52.5%	<i>celebridade</i> (celebrity), <i>instabilidade</i> (instability), <i>modernidade</i> (modernity)
-ia	758	21.0%	<i>valentia</i> (boldness), <i>alegria</i> (joy) , <i>acefalia</i> (acephalia)
-ismo	277	7,7%	<i>brilhantismo</i> (brightness), <i>heroísmo</i> (heroism), <i>realismo</i> (realism)
-ice	230	6.4%	<i>meiguice</i> (gentleness), <i>parvoíce</i> (silliness), <i>velhice</i> (old age)
-ez	154	4.3%	<i>estupidez</i> (stupidity), <i>rapidez</i> (rapidity), <i>cupidez</i> (cupidity)
-eza	128	3.6%	<i>esperteza</i> (cleverness), <i>firmeza</i> (firmness), <i>vagueza</i> (vagueness)
-idão	86	2.4%	<i>gratidão</i> (gratitude), <i>lentidão</i> (slowness), <i>vastidão</i> (vastness)
-ura	79	2.2%	<i>doçura</i> (sweetness), <i>espessura</i> (thickness), <i>loucura</i> (madness)

Table 3: Relative productivity of DPN suffixes in Portuguese

#### 4.4. Adjectival Polysemy

For the deverbal event nouns in Portuguese, the meaning of the noun is, of course, dependent on the meaning of the verb. And if the verb is polysemous, so is the derived event noun. But if the verb *acabar* (finish; conclude; destroy; end up..) is polysemous,

then *acabamento* is the action or effect of the verb in each of these meanings (Janssen & Ferreira 2007).

For DPN, this does not hold. There are many cases where the PN only refers to the state or quality related to a specific meaning of the adjective. And when there are various meaning for the adjective, there often is a different DPN for the different adjectival word-senses. Examples are the following:

(a) *fino* (thin; delicate);

*finura* (the fact of being delicate);

*fineza* (the fact of being thin).

(b) *bravo* (angry; brave);

*braveza* (the fact of being angry);

*bravura* (the fact of being brave)

Because the meaning of the DPN is dependent on a specific meaning of the base adjective, the suffix should be said not to apply to the adjective itself, as in the case of deverbal nouns, but to a word-sense of the base adjective. That means that all the rules about the applicability of suffixes to adjectives have to be related to word-senses of the adjectives, rather than to adjectives themselves. In the analyses that follow, wherever an adjective is mentioned, it should therefore always be interpreted as an adjective in a specific meaning, even if this is not explicitly mentioned.

#### 4.5. Adjective-bases – Semantic Restrictions

In this section we present the result of the preliminary analysis of the restrictions, based on the sample set of 40 DPN selected at random from the database for the least productive of the suffixes (*-ice*, *-ez*, *-eza*, *-idão*, and *-ura*), when compared with the semantic type of the base adjective. The data presented here relate only to the sample

sets of 40 adjectives, not to the full set of all the DPN in the database.

The selected adjective with the suffix *-ice* show that this suffix is not sensitive to the morphological structure of the adjective, combining with both morphologically simple and complex adjective. However, the suffix is sensitive to the semantics of the base adjective: it only combines with adjectives that identify human qualities, which are mainly behavioural, but also qualities related to social stereotypes, age-indication and human skills, as shown in table 4.

Human Behaviour	29	<i>guloso</i> (greedy) > <i>gulosice</i>
Social Stereotype	6	<i>saloio</i> (redneck) > <i>saloice</i>
Age-Indication	3	<i>veterano</i> (veteran) > <i>veteranice</i>
Human Skills	2	<i>besta</i> (beast) > <i>bestice</i>

Table 4: Semantic values of adjective-bases from suffix *-ice*

The suffix *-ez* only applies to morphological simple Latin roots. From set of nouns selected for the suffix *-ez* a large number have base adjective describing physical states, and also a considerable number of human psychological states, as shown in Table 5. Although these two groups are rather different types of properties, there is a correlation between them in the lexicon: many of the adjectives characterizing psychological behaviour are metaphorical extensions of adjectives referring to a physical state. For instance, the adjective *frigido* (frigid) can mean either something really cold, or a psychological state, meaning *insensible* or *indifferent* (as is the case with the English adjective *frigid*).

Physical States	22	<i>sólido</i> (solid) > <i>solidez</i>
Human Psychological States	14	<i>estúpido</i> (stupid) > <i>estupidez</i>
Human Physical States	4	<i>surdo</i> (deaf) > <i>surdez</i>

Table 5: Semantic values of adjective-bases from suffix *-ez*

The nouns ending in *-eza* have a preference for adjectives which indicate physical properties and human behaviour types (cf. table 5). In many cases, this suffix co-occurs with the suffix *-ez*, especially for those adjectives which can be used to define a physical property or a human behaviour. An example is *esquivo* (elusive - either something difficult to catch or a person who is always escaping from something), which gives rise to both *esquiveza* and *esquivez*.

Physical Properties	16	<i>fraco</i> (weak) > <i>fraqueza</i>
Human Behaviour	12	<i>gentil</i> (gentle) > <i>gentileza</i>
Qualification	7	<i>certo</i> (sure) > <i>certeza</i>
Dimension	5	<i>baixo</i> (small) > <i>baixeza</i>

Table 6: Semantic values of adjective-bases from suffix *-eza*

The suffix *-idão* also only applies to simple base. However, there is a larger variety in the semantic types of the base adjectives, as shown in table 7, including physical properties, colour adjectives, and several others.

Physical Properties	19	<i>escuro</i> (dark) > <i>escuridão</i>
Colours	16	<i>vermelho</i> (red) > <i>vermelhidão</i>
Qualification	3	<i>apto</i> (able) > <i>aptidão</i>
Others	12	<i>lento</i> (slow) > <i>lentidão</i>

Table 7: Semantic values of adjective-bases from suffix *-idão*

The suffix *-ura* cannot combine with complex adjectives either, and largely selects physical properties of human behaviour adjectives, as can be seen in table 7. There is a large correlation between the suffixes *-idão* and *-ura*: all the DPN for colour ending in *-idão* also have a corresponding DPN on *-ura*: *negro* (black) gives rise to *negrura* and *negridão*.

Physical Properties	19	<i>doce</i> (sweet) > <i>doçura</i>
Human Behaviour	16	<i>bravo</i> (brave) > <i>bravura</i>
Colour	3	<i>branco</i> (white) > <i>brancura</i>

Table 7: Semantic values of adjective-bases from suffix *-ura*



Although this is only a very small and preliminary analysis, it shows most of the DPN with the less productive suffixes shown here have an adjectival base denoting physical properties, and that some of these property nouns often have competing DPN with other suffixes as well, which have a figurative definition related to human behaviour.

The fact that *-ez*, *-eza*, *-idão* and *-ura* only select morphologically simple bases may explain their weak productivity when compared with that of suffixes like *-idade*, *-ia*, *-ismo*, and *-ice*, cf. table 3.

## 5. Conclusions

In this paper, we have shown the method of construction and analysis of the database of all lexicalized de-adjectival property nouns in Portuguese. The database of DPN represent the structure of the property nouns in a more structure manner, since dictionaries have a diverse and often non-systematic treatment of the property nouns, with too much of a heterogeneous treatment in the etymology for a structural linguistic analysis based directly on the dictionary. The manual verification applied in the construction of the database was necessary in order to create a more uniform treatment.

Although the data of the analysis presented here are very preliminary, they show that there is a relation, although not always clear at first sight, between morphological structure, meaning and referential possibilities, which can be uncovered by a data-driven analysis of DPN. We have confirmed that when for one adjectival base it is possible to build different nouns, these nouns have in principle different denominative capacities. We have also shown that the types of adjectives to which the less productive suffixes can apply is often rather limited. With further research, both on the process of DPN formation and on the other derivational processes in Portuguese, we hope to present these restriction and other restrictions in more detail.

## References

- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. 2 vols. Tübingen: Max Niemeyer Verlag.
- Corbin, D. (1991). "Introduction – La formation des mots: structures et interprétations". *Lexique*, vol. 10, 7-30.
- Corbin, D. (1997). "La représentation d'une «famille» de mots dans le Dictionnaire dérivationnel du français et ses corrélats théoriques, méthodologiques et descriptifs". *Recherches linguistiques de Vincennes*, 5-37.
- Correia, M. (1994). *Bases digitais na União europeia*. Unpublished paper presented at Simpósio de Lexicologia, Lexicografia e Terminologia (Araraquara, Brazil).
- Correia, M. (1999). *A denominação das qualidades: contributo para a compreensão da estrutura do léxico português*. Dissertation. Lisbon: Universidade de Lisboa.
- GDLP (2004). *Grande Dicionário Língua Portuguesa*, 1a ed., Porto: Porto Editora.
- Houaiss. *Dicionário Eletrônico Houaiss*.
- DLPC. (2001). *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*, Lisboa: Academia das Ciências de Lisboa / Editorial Verbo.
- Hearst, M. 1998. "Automated discovery of wordnet relations". In: Fellbaum (ed.) *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Janssen, M & J. P. Ferreira. (2007). "Nomes Deverbais Eventos, constucção de um novo dicionário". Paper presented at the XIII Encontro do APL, Évora, Portugal.
- Janssen, M. (2007). "Portal da Língua Portuguesa". *Procesamiento del Lenguaje Natural*. vol. 39, 297-298
- Michaelis. *Dicionário Multimídia Michaelis 5.0*.
- Rio-Torto, G. M. (2006). "Para uma gramática do adjetivo". *Alfa*, vol. 50 (2), 103-129.