An Overview on Portuguese Nominalizations

Livy Real[1] and Alexandre Rademaker[2]

[1] IBM Research (Brazil)
[2] IBM Research - FGV/EMAp (Brazil)

Abstract. We discuss nominalizations in Portuguese formed by the suffix *-ura*. We have done a corpus-based description of the behavior of these nominals and proposed a type ontology to categorize them. In order to offer a rich description, we also tested all words formed by *-ura* in co-predication contexts to check if their types could be co-predicated. Although our main goal was to produce a corpus-based description on those nominals, we have found that may be the frequency of use of a given word has a special role on the acceptability of co-predication between different senses of a nominalization.

1 Introduction

If we consider the last decades of formal linguistic studies, the behavior of nominalizations is a very recurring topic, specially because it imposes challenges to lexical semantics theories, knowledge representation systems, and other areas of formalism.

Generally, studies on nominalizations consider only prototypical nominals (as *construction*, *destruction* and *translation*) in the search for generalizations on their behavior or even while deciding what is the best way to represent (or to understand) them. This research intents to look extensively into one specific kind of nominalization: deverbal nouns formed by suffix *-ura* in Brazilian Portuguese (henceforth BP) as *abertura* 'opening', *assinatura* 'signing/signature' and *brancura* 'whiteness'. We hope to reach an enriched description of a relevant fragment of deverbal nominals in BP, considering all possible senses of each noun and possible co-predications between them.

2 Our motivations

We focus on BP because it is not a very mainstream language used in linguistic studies, even though it has similarities with most neo-Latin languages. The choice of the suffix *-ura* was made for two reasons: we already know [5] how this suffix morphologically works and, more than that, the discussion of their behavior in BP probably could be used to understand other suffixes in other neo-Latin languages (as *-ura* in Catalan or *-ure* in French) as they have similar suffixes working on deverbal nominalization processing.

The decision to work on nominalizations formed by a specific suffix was because we want to do a corpus-based overview of one fragment of nominalizations

in order to include action nominals — deverbal nouns which carry eventive readings, cf. [9] — and also non-prototypical nominalizations, that is, nominalizations whose main value is not an eventive or processual meaning.

We hope that our investigation on *-ura* nominals can hint us on the behavior of similar nominalizations in other languages, since *-ura* is a morpheme which has at least eight meanings cataloged by literature on Portuguese ([8, 7, 4]), which include eventive, resultative, locative and collective readings, just as other well-known morphemes like *-ung* in German and *-age* in French. In those lines, we expect that working on these nominal will get new perceptions on what is going on nominalizations.

3 Methodology

To produce an empirical description of this data, we have chosen to work on a corpus-based analysis. All of our descriptions are based on written corpora and the test sentences were checked in a given context with at least three speakers not related to linguistics studies. We believe that it is important to enlarge the discussion on nominalizations since many researches take into account just very well known and simple sentences and constructions.

All the nouns used in the research came from OpenWordNet-PT [3].[3] We extracted from the list all nominals finished by the graphic form "ura" (442 words) and manually selected the nouns formed by the suffix *-ura* (150 words). It is important to note that, as we extracted synsets from OpenWordnet-PT, some words were counted more than once, as they appear in more than one synset.

Then, to categorize the selected nominals, we have checked Porto's Dictionaries [4], Caldas Aulete Dictionary [5] and Houaiss Dicionary [6] for all the possible meanings to each noun in our list. We also checked the presence of each noun on Corpus Brasileiro (developed by Tony Berber Sardinha & alii) trough AC/DC tool [2] [7] which has more than 1 billion words extracted from various textual genders. Google engine was used when Corpus Brasileiro was not enough.

In addition we have relied in recent literature on nominalizations and BP to capture some more insight about the typology those nouns could assume. Almost all the sentences analyzed in this work comes from Corpus Brasileiro and some of them were found on Google search engine in different trustworthy websites.

---

[3] http://wnpt.brlcloud.com/wn/

[4] http://www.infopedia.pt

[5] http://www.aulete.com.br

[6] http://www.houaiss.uol.com.br

[7] http://www.linguateca.pt/ACDC

4 Analysis

Earlier studies have shown that eventive nominalizations in Portuguese can assume at least the following readings (cf. [4]): event, result, physical result, resultative state, abstract result, locative and collective.

Following this categorization, we checked on dictionaries and confirmed on corpora the possible meanings to each noun from our list. From that analysis, we found that nominals formed by *-ura* can have the following readings: event, result, physical result, locative, collective, means, property, instrument, a given portion, rest, function, duration of a function, science/art, as shown by the examples bellow.

Event Deduziu-se que a mãe lhe deu muita chicotada a cada **travessura**. 'It was deduced that the mother gave him a lot of whiplashes at every **trick** (every time he misbehaved).'

Result A análise do material revelou que, 30 dias após a microenxertia, ocorreu a **soldadura** parcial dos microenxertos. 'The analysis of the material showed that, 30 days after micrografting, occurred the partial **welding** of micrografts.'

Physical Result A **varredura** mostra somente picos, como pode ser visto na Figura 8, onde o espelho de simetria de 0° é mostrado. 'The **scan** shows only peaks, as it can be seen in Figure 8, where the symmetry mirror of 0° is shown.'

Locative Meu certificado está na pasta com meus documentos na **prefeitura**, mas o prefeito não o reconheceu. 'My certificate is the folder with my documents in the **city hall**, but the Mayor did not recognized it.'

Collective Uffizi tem o mais completo testemunho do século XV, um momento decisivo da história da arte, marcado pela passagem da tradição bizantina medieval para a **pintura** do Renascimento. 'Uffizi has the most complete reference of XV Century, a decisive moment of Art History, marked by the passage of Medieval Bizantine tradition to the Renaissance **painting**'

Means A narrativa é um cavalo: um meio de transporte cujo tipo de **andadura**, trote ou galope, depende do percurso a ser executado. 'The narrative is a horse: a means of transportation which type of **gait**, trot or gallop, depends on the route to run.'

Property Possui cerca de 48% de umidade e 24% de **gordura**. 'It has around 48% of umidity and 24% of **fat**.'

Instrument Caricaturizada, a gostosona desfila engravatada, com chapéu, **abotoadura** e tudo mais. 'Caricatured, the hot girl parades with tie, hat, **cufflink** and everything.'

A given portion Assim verificamos que os 587 pés que aquelas dez propriedades dos Calça Pereira comportam podiam render uma média de 23,5 **moeduras**, isto é, uns 940 alqueires de azeite, que valeriam, ao preço de 60 reais o alqueire, 5640 reais. 'Thus we have verified that the 587 feet of those ten properties from Calça Pereira family include could yield an average of 23.5 **milling portions**, that is, some 940 acres of olive, which would be worth at the price of 60 reais per bushel, 5640 reais.'

Rest O arroz-caril, confeccionado com especiarias e **moedura** de coco, era característico de Goa e estava muito difundido em Moçambique.[8] 'The rice-curry, made with spices and coconut **grinding**, was characteristic of Goa and was widespread in Mozambique.'

Function Mario renunciou à **magistratura** em novembro. 'Mario resigned to the **magistracy** in November.'

Duration of a function Para a **legislatura** de 1995-1998, os dados provêm do Brasil. 'For the **legislative period** 1995-1998, the data comes from Brazil.'

Science/Art A Itália exprimiu-se, durante certos séculos, pela **arquitetura, escultura, pintura**. 'Italy expressed herself, during some centuries, by the **architecture, sculpture, painting**.'

We also cataloged all the lexicalized values, which are not closely related to the verb base or that cannot be found recurrently in the lexicon. From our 150 nouns list, 33 presented a lexicalized and idiosyncratic meaning. It is worth knowing that two possible types of action nominal formed by -*ura* are not possible (or frequent) in BP: resultative state and abstract result.

Within this categorization, we looked for possible generalizations on the behavior of those nominals and also specific cases that can confirm (or not) the patterns brought by the literature.

After this categorization, we searched for co-predications within the nouns on Corpus Brasiliero and Google. For our surprise, we did not find any co-predication among the nouns that are not used frequently in BP, we only found co-predications among very commonly used nouns (as *assinatura* 'signature' or 'signing' and *abertura* 'opening').

After that, we produced some sentences combining different types of nouns of frequent and not so frequent use to discover what kind, if any kind, of co-predication would be possible. We have tested all these sentences in context with at least three native speakers of BP with no knowledge of linguistic study theories. Surprisingly again, all the words which are not very commonly used in BP do not accept co-predication in any situation, even when the tested types were 'result' and 'event':

Apesar de ter durado uma hora, a abertura foi proveitosa a todos os alunos. 'Although it lasted one hour, the opening was beneficial to all students.'

*Apesar de ter durado uma hora, a brochura deixou os livros lindos. 'Although it lasted one hour, the brochure left the books beautiful.'

*A mordedura foi rápida, mas deixou uma cicatriz. 'The bite was quick, but left a scar.'

The examples above show co-predication attempts with the words *abertura*, *brochura* and *mordedura* which have a very different number of occurrence in Corpus Brasileiro: *abertura* occurs 70699 times, *brochura* appears 371 times and *mordedura* has only 120 tokens. Therefore, it seems that the frequency of use of a given noun has some influence on its co-predication acceptability, since *abertura* and *mordedura* have the exact same type structure and their verb bases, *abrir* 'open' and *morder* 'bite', have a very similar behavior.

---

[8] http://www.scielo.br/pdf/hcsm/v21n2/0104-5970-hcsm-21-2-0609.pdf

From all this, the generalizations we could get are related to the type structure assumed:

1. A nominal form that has the type 'rest' belongs to the type 'event' (as *lavadura* 'washing' and *varredura* 'scan'), but co-predications between them are impossible;
2. A noun that belongs to the type 'a given portion' (as *moedura* 'milling' and *semeadura* 'sowing') has always the following types: 'event', 'result', 'event.result', but any co-predication with 'a given portion' is blocked;
3. Every noun that belongs to 'duration of a function' also holds the same 'function' type;
4. Nouns that belong to the type 'means' do not belong in any other type;
5. All lexicalized senses can not be co-predicated with any other type.

5   Conclusions and future work

From this small experiment, we can conclude that some nominalizations have a very strict behavior and some others are much more flexible. It seems that another linguistic phenomena could have an influence on this flexibility: the (high or low) frequency that a word is used may be the key to its co-predication acceptability. Besides, our generalizations above are about how particular nominals behave in BP. Perhaps our main contribution with this small analysis is to call the attention to the fact that ordinary nominalizations (highly used words) have the tendency to allow co-predication between their types. Meanwhile uncommon words, with the same type structure and the very same morphological formation, do not.

For this note, we do not present a formalization of our results, but, as argued by [4], it seems that a tool like Montagovian Generative Lexicon is able to deal with this very idiosyncratic behavior of nominalizations formed by *-ura*, as it has a very flexible mechanism to describe (in)felicitous co-predication. For now, the application of Montagovian Generative Lexicon on *-ura* nominalizations and a deeper investigation on the relation between felicitous co-predication and the frequency of use of a given noun remains for future analysis.

References

1. REAL, L. RETORÉ C. *Deverbal Semantics and the Montagovian Generative Lexicon Λ Tyn.* Journal of Logic, Language and Information, v. 1, p. 1, 201, 2014.
2. SANTOS, D. SARMENTO, L. *O projecto AC/DC: acesso a corpora/disponibilização de corpora*, APL, pp. 705-717, Porto, 2002.
3. de Paiva, V. RADEMAKER, A. MELO, G. *OpenWordNet-PT: An Open Brazilian WordNet for Reasoning*, Proceedings of the 24th International Conference on Computational Linguistics, 2012.
4. REAL, L. *Nominalizações.* Ph.D thesis, Universidade Federal do Paraná, Curitiba, Brazil, 2014.

5. REAL, L. *Morfologia Categorial.* Undergraduation final work, Universidade Federal do Paraná, Curitiba, Brazil, 2006.

6. BASSAC, C. MERY, B. RETORÉ, C. *A Montagovian generative lexicon* in CSLI, Formal Grammar 2007.

7. ROCHA, L. C. *A nominalização no português do Brasil.* Revista de Estudos da Linguagem, 8 (17), 1999.

8. SANDMANN, A. J. *Formação de palavras no português brasileiro contemporâneo.* Curitiba: Scentia et Labor: Ícone, 1988.

9. MELLONI, C. *Polysemy in word formation: the case of deverbal nominals.* University of Verona: Dissertation, 2007