

Embedding NomLex-BR nominalizations into OpenWordnet-PT

Livy Maria Real Coelho¹ Alexandre Rademaker^{2,5}
Valeria de Paiva³ Gerard de Melo⁴

UFP

IBM Research

Nuance Comms.

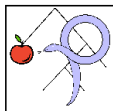
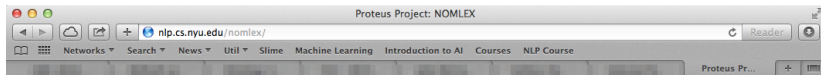
Tsinghua University

FGV/EMAp

February 1, 2014



The English NomLex



Nomlex

[Proteus Project](#)

[Department of Computer Science](#)
[New York University](#)

Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barrett and Ruth Reeves

Overview

NOMLEX (NOMinalization Lexicon) is a dictionary of English nominalizations developed by the Proteus Project at New York University under the direction of Catherine Macleod. NOMLEX seeks not only to describe the allowed complements for a nominalization, but also to relate the nominal complements to the arguments of the corresponding verb. The complements of the nominalization are described in terms of the COMLEX Syntax verb subcategorization patterns of its associated verb. See the COMLEX Syntax Manual for information on the verbal complements. We identify both the main verbal arguments (subject, direct object, and indirect object), which may map into a variety of nominal positions, and the oblique verbal complements, which map more directly into nominal complements. The argument correspondences are specified through a combination of explicit information in the lexical entries and general linguistic constraints on the correspondences. We have 1025 entries of several types of lexical nominalizations, including over 1000 distinct words. These words were selected from lists of frequently appearing nominalizations in our corpus (which includes Brown and the Wall Street Journal). There are multiple entries for certain graphonyms. For example, "deduction" has two NOMLEX entries ("deduction1" and "deduction2"), one corresponding to the verb "deduct" and the other corresponding to the verb "deduce". We released the alpha version of NOMLEX on January 15, 1999, a small update to the alpha version (the Alpha 2 version) on March 12, 1999. Finally, the 2001 version was released on October 19, 2001. The latter is downloadable from this website (see below) and freely available for use by all. We would appreciate feedback from this usage to macleod@cs.nyu.edu.

A sample entry follows:

```
(nom      :orth "promotion"
          :verb "promote"
          :nom-type((verb-nom))
          :verb-subj ((n-n-mod) (det-poss))
          :verb-subc ((nom-np :object ((det-poss) (n-n-mod) (pp-of)))
                     (nom-np-as-np :object ((det-poss) (pp-of)))
                     (nom-possing :nom-subc ((p-possing :pval ("of"))))
                     (nom-np-pp :object ((det-poss) (n-n-mod) (pp-of))
                                :pval ("into" "from" "for" "to"))
                     (nom-np-pp-pp :object ((det-poss) (n-n-mod) (pp-of))
                                :pval ("for" "into" "to") :pval2 ("from"))))
```

NomLex (cont.)



Alexander's destruction of the city happened in 330 BC.

- ▶ a dictionary of English nominalizations, under Catherine Macleod.
- ▶ relate the nominal complements to the arguments of the corresponding verb.
- ▶ 1025 entries of several types of lexical nominalizations.
- ▶ first version on January 15, 1999, latest version October 2001 downloadable from <http://bit.ly/1aZWQmh>

Nomlex (cont.)

```
(nom : orth "promotion"  
      : verb "promote"  
      : nom-type ((verb-nom))  
      : verb-subj ((n-n-mod) (det-poss))  
      : verb-subc ((nom-np : object ((det-poss)(n-n-mod)(pp-of)))  
                    (nom-np-as-np : object ((det-poss) (pp-of)))  
                    (nom-possing : nom-subc ((p-possing : pval ("of")))))  
                    (nom-np-pp : object ((det-poss) (n-n-mod) (pp-of))  
                                          : pval ("into" "from" "for" "to"))  
                    (nom-np-pp-pp : object ((det-poss) (n-n-mod) (pp-of))  
                                           : pval ("for" "into" "to") : pval2 ("from")))))
```

Related Works

- ▶ Nominalizations have been studied for more than 4 decades (Chomsky, 1970).
- ▶ NomLex-Plus (Meyers et al., 2004). Extension of NomLex with 7.050 nominalizations.
- ▶ The NomBank Project (Meyer, 2007) <http://bit.ly/1d5G7L9>.
“ mark the sets of arguments that co-occur with nouns in the PropBank Corpus, just as PropBank records such information for verbs... firmly on the shoulders of NOMLEX...”
- ▶ Berkeley FrameNet (<https://framenet.icsi.berkeley.edu/>).
11600 lexical units based on frame semantics supported by corpus evidence. Deverbal nominalizations are annotated as events (in the frame of verbs) or entities/results (diff. semantic frame).
- ▶ FrameNet-Brazil, <http://www.ufjf.br/framenetbr/>.

Using for NLP (IE)

- ▶ To write maps between IE patterns for active clauses to IE patterns for nominalizations.
- ▶ Active clause: “IBM appointed Alice Smith as vice president”.
- ▶ Passive clause: “IBM’s appointment of Alice Smith as vice president” and “Alice Smith’s appointment as vice president”.

Main use for NLP (IE) (cont.)

The Proteus Extraction System starts with:

np(C-company) vg(appoint) np(C-person) "as" np(C-position)

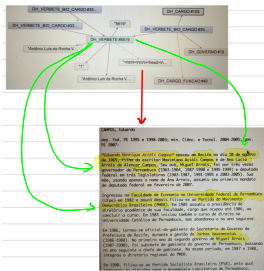
Meta rules to produce passive clause pattern:

np(C-person) vg-pass(appoint) "as" np(C-position) "by"
np(C-company)

When a pattern matches the input, the pieces corresponding to its constituents are used to build a semantic representation of the patter (e.g. logical form).

vg = verb group (plus auxiliaries). vg-pass = passive verb group.

Project Motivation: DHBB



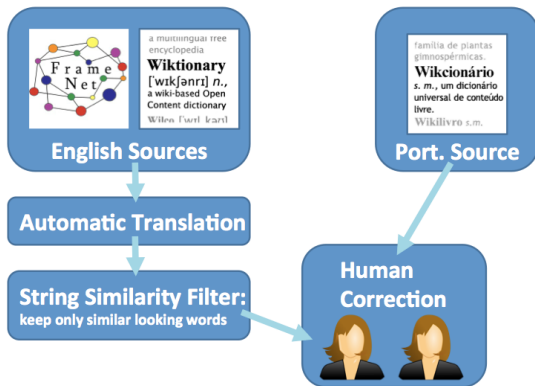
- ▶ 7.5K entries Brazilian Historical Biographic Dictionary (DHBB).
- ▶ Enrich the structure (semantics). Uniform data treatment (standards and interlinks between collections).
- ▶ NLP of DHBB entries: (1) word sense disambiguation with openWordnet-PT; and (2) named entity recognition to make links. (133K proper names)

We need grammars, lexical resources, ontologies, KBs, automated theorem provers etc to reason about knowledge extracted from text. This will empower QA, KE, MT, personal assistants and other systems.

Nominalizations in Portuguese

- ▶ Nominalizations: difficult to deal with in KR systems, harder to obtain the arguments of nominal predicate;
- ▶ NOMLEX project (Macleod et al., 1998) provides a well-established, open access baseline;
- ▶ nominalizations with the suffixes **-ção**/*-ion*, **-mento**/*-ment* and **-or**/*-er*, which work well in Portuguese;
- ▶ E.g. construção (*construction*), adiamento (*adjournment*) and escritor (*writer*);
- ▶ 90% of the original resource **easily** manually translated.

How we expanded it



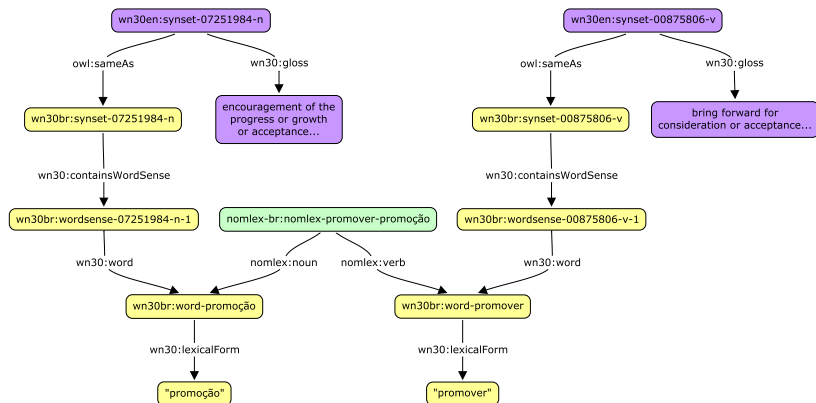
We **translate** both noun/verb by looking up in extractions from the EN and PT Wiktionary dumps, generating all combination of noun/verb translations. **Filter** to compare the noun and verb translations to see if they are similar enough to be morphologically related. Other experiments with DHBB and openWordnet-PT.

NomLex-BR

- ▶ a dictionary of Portuguese nominalizations
- ▶ Relate nominals to corresponding verbs
- ▶ Over 2,539 entries of several types of lexical nominalizations
- ▶ first version of NOMLEX-BR in 2011, much expanded 2013
- ▶ Freely available for download and embeded in openWordnet-PT.
- ▶ A RDF vocabulary to describe nominalizations. Future extensions to cover more information from COMLEX and COMNOM (extension from NomBank).
- ▶ URI for the schema,
`http://arademaker.github.com/nomlex/schema/!` Need a better and stable URI.

“Construção da rodovia Transamazônica, na década de 70, pelo governo Medici, uma das obras faraônicas da ditadura militar.”

Embedding in openWordnet-PT



But **nomlex:noun** and **nomlex:verb** should point to **wn30:WordSense** not **wn30:Word**! Future work!

By Provenance

See <http://bit.ly/Mohmni>

```
select ?prov (count(?x) as ?total) {  
  ?x a nomlex:Nominalization ;  
      dc:provenance ?prov .  
}  
group by ?prov
```

provenance	total
nomlex	1032
wiktionary-pt	61
wiktionary-en	91
framenet	142
nomage	262
dhbb	159
openWordnet-PT	82
linguateca	484

By suffix

See:

- ▶ <http://bit.ly/LmAXn4>; and
- ▶ <http://bit.ly/1fKErKr>.

Result:

suffix	total
mento	329
ção	660
or	891

Some other cases <http://bit.ly/1fyia3a>.

Results

- ▶ Extension of OpenWN-PT aims at incorporating links to connect deverbal nouns with their corresponding verbs.
- ▶ The integration into OpenWN-PT will facilitate their use for linguistic research as well as information extraction
- ▶ Incorporating NOMLEX-BR data into OpenWN-PT has shown itself useful in pinpointing some issues with the coherence and richness of OpenWN-PT.
- ▶ the word **abasement** corresponds in NOMLEX to the verb **abase**, and thus we would like a similar correspondence between the Portuguese noun “aviltamento” and the verb “aviltar” (suggested translations). OpenWN-PT simply has two synsets “humilhar, abaixar” and “humilhar, rebaixar”. The more common verb humilhar is repeated, while the uncommon aviltar was left out.

Next Steps

- ▶ Finish to embed Nomlex-BR into OpenWN-PT (anchor floating words, <http://bit.ly/1aQdpkr>).
- ▶ Work with Claudia Freitas and Hugo Gonçalves on leveraging Linguatecas PAPEL, Cartão, ACDC and Floresta Sintá(c)tica.
- ▶ Lists from Linguateca's resources complement NomLex-BR using corpora and make sure our resource is not simply a translation.
- ▶ Adding the Portuguese terms that satisfy different relations? OpenVerbNet-PT? Glosses? Classification of nominalizations?
- ▶ We are developing our own web interface for browsing and collaborative editing. Most important pending issue!
- ▶ Use and test the accuracy of the resource! More applications!

Conclusion



- ▶ We presented NomLex-BR, an lexicon of nominalizations in Brazilian Portuguese.
- ▶ NomLex-BR is embedded into OpenWordNet-PT and shares its RDF representation.
- ▶ Recent improvements include better coverage: newer suffixes and Nomage incorporation.
- ▶ The work with Nomlex-BR helped us to improve openWordnet-PT (new words, senses).

The data is freely available from

<http://github.com/arademaker/wordnet-br/> and a SPARQL

Endpoint at <http://logics.emap.fgv.br:10035>.

Obrigado!

Synset 01146493-a

Danish	<i>taknemmelig</i>
English	<i>thankful, grateful</i>
Finnish	<i>kiitollinen</i>
French	<i>reconnaissant</i>
Galician	<i>grato, agradecido</i>
Indonesian	<i>bersyukur, berterima kasih, tanda terima kasih, terhutang budi</i>
Italian	<i>grato, riconoscente</i>
Japanese	忝い, 有り難い, 感謝を感じた, 幸甚, ありがたい, 有難い, 感謝を表した
Bokmål	<i>takknemlig</i>
Portuguese	<i>reconhecido, grato, agradecido</i>
Thai	ซึ่งสำนึกในบุญคุณ
Malaysian	<i>bersyukur, berterima kasih, tanda terima kasih, menampakkan tanda kesyukuran, memperlihatkan tanda kesyukuran, terhutang budi</i>

Eng: feeling or showing gratitude; "a grateful heart"; "grateful for the tree's shade"; "a thankful smile";