AN ENGLISH-PORTUGUESE PARALLEL CORPUS FOR THE STUDY ON THE SEMANTICS OF NOUN-NOUN COMPOUNDS

Lilian Figueiredo Teixeira • Universidade de Vale do Rio dos Sinos – UNISINOS
Rui Leiva de Oliveira Chibum • Universidade – Pq CNPq

INTRODUCTION
This study suggests a semantic analysis of the noun–noun compounds in English in a magazine context, presenting an initial reflection on the translation correspondents.

THEORIES AND CONCEPTS
• Noun–noun compounds: two nouns (NN) in which there is a pre–modifier followed by a head noun (BARKER; SIZA-PAYNE, 1998).
• Parallel Corpus: each line of an original text is aligned with its translation in the second language (BERGER SARDONIO, 2004).
• Translations metaphorical: part of a translational unit, and the textual context has no to be taken into account (HERRANZ, 1994).
• Semantic puzzle: Ryder (1994) suggests a classification based on linguistic templates, which are schemas that include semantic characteristics of the components and of the compound structures.
• Frames: provide a description of a specific context, through the identification of related actors and lexical units (FILLMORE, 1982).

OBJECTIVE
To identify how the semantics of the compounds could suggest some predictability on the structure of the noun compounds in the target language.

METHODOLOGY
• Ten editions of the National Geographic magazine – English and Portuguese, from August 2007 to May 2008.
• Itemization and morphological annotation of the articles in English (TreeTagger).
• An extractor was used, a tool that provides a list of possible compounds, by Lucian Lennox.
• Then some semantic patterns were identified (RYDER, 1994) – Wordworth Tool for the compound uses.
• The two texts were machine tagged (Vanilla Alpino) and the translation equivalents were identified.

RESULTS
Most frequent semantic relations

<table>
<thead>
<tr>
<th>N. OF OCCURRENCES</th>
<th>RELATION</th>
<th>ENGLISH</th>
<th>PORTUGUESE</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>IS LOCATED IN</td>
<td>school play</td>
<td>peça escolar</td>
</tr>
<tr>
<td>37</td>
<td>HAS CONSTITUENT PART</td>
<td>church floor</td>
<td>solo da igreja</td>
</tr>
<tr>
<td>25</td>
<td>SERVES TO</td>
<td>car keys</td>
<td>chave do carro</td>
</tr>
<tr>
<td>9</td>
<td>COMES FROM</td>
<td>cane juice</td>
<td>calda do cauão</td>
</tr>
<tr>
<td>9</td>
<td>HAPPENS IN TIME</td>
<td>night school</td>
<td>escola noturna</td>
</tr>
</tbody>
</table>

• 81% of the 185 NN compounds translated present the structure formed by nouns, preposition "de" and nouns. Other structures: "in adjective", "in N", "in N as N", "of N as N", "of N as N of N" and "in N as N of N". Sometimes one noun in Portuguese. "SERVES TO" many translation equivalents are formed by noun and adjective, which sounds much more natural than a NN construction.

The authors refer to their houses because, pigs, farm tools, grains, and the wolves clothes and silver heirlooms of grandparents and mothers.

• On agrícolas avulam nozes pretimigens cumo, nozes, forrageiras agrícolas, depósitos de galês, as roças tecidas e as heranças de prata de seus avós e mães.

• Ethanol and biodiesel are now made from food crops like corn and soybeans, but in principle any plant material will do. O etanol e o biodiesel, produzidos a partir de cereais alimentícios como milho e soja, em princípio qualquer ser feito com qualquer vegetal.

• El es ost de uniforme escolar, diz Shavzieh. And students don’t need space suits, radiation shields, tolls, exercise busses, a bail-out system during launch, or any consumables to speak of except energy.

CONCLUDING REMARKS
The preposition "de" in Portuguese can represent several semantic relations and deserves more studies on its polysemy. An important aspect on the translation is the cultural influence on the translator options. Contributions to improve the language processing tasks, such as machine translation systems.

REFERENCES