PAPEL: a lexical ontology for Portuguese

Hugo Gonçalo Oliveira, Paulo Gomes, Nuno Seco
Linguateca, node of Coimbra, DEI-FCTUC, CISUC

Diana Santos
Linguateca, node of Oslo, SINTEF ICT
Contents

- Project objectives
- MRDs (as a source of semantic knowledge)
- Similar resources
- Building PAPEL
- Detailed example
- Current precision
- Further work/possible directions
Project objectives

PAPEL stands for Palavras Associadas Porto Editora Linguateca
MRDs as a source of semantic knowledge

Since 1970's (Calzolari 1977, Amsler 1981, ...)
Restricted and predictable vocabulary

Typical definition structure:

- **genus**: the superordinate concept
- **differentia**: properties for distinction between instances of the same superordinate concept.
Similar resources

Princeton WordNet (Miller 1990)
- freely available
- widely used by NLP researchers
- created “from scratch”
- synsets + relations
- different kinds of relations for words of different grammatical categories
- Hypernymy/Hyponymy, Meronymy/Holonymy.
Similar resources

Microsoft's MindNet (Richardson et al. 1998)

- automatically created with tools to extract relations from MRDs
- a large amount of relations: Attributive modification, Causation, Classifier, Goal, Hypernymy, Intensifier, Location, Manner, Means...
- a web interface: MNEX
  (http://atom.research.microsoft.com/mnex/)
Similar resources

FrameNet (Baker et al. 1998)
- description of semantic frames

WordNet.BR (Dias da Silva 2004)
- as far as we know only contains the Synonymy and Antonymy relations.

WordNet.PT (Marrafa et al. 2006)
- little information shared
- website/search interface has not been working for the past year...
Examples of relations to include:

<table>
<thead>
<tr>
<th>Relation</th>
<th>Read</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIPERONIMO-DE(X, Y)</td>
<td>X is a HIPERONIMO-DE Y</td>
<td>HIPERONIMO-DE(animal, cão)</td>
</tr>
<tr>
<td>HIPONIMO-DE(X, Y)</td>
<td>X is a HIPONIMO-DE Y</td>
<td>HIPONIMO(cão, animal)</td>
</tr>
<tr>
<td>CAUSADOR-DE(X, Y)</td>
<td>X is a CAUSADOR-DE Y</td>
<td>CAUSADOR-DE(vírus, doença)</td>
</tr>
<tr>
<td>RESULTADO-DE(X, Y)</td>
<td>X is a RESULTADO-DE Y</td>
<td>RESULTADO-DE(doença, vírus)</td>
</tr>
<tr>
<td>MEIO-PARA(X, Y)</td>
<td>X is a MEIO-PARA Y</td>
<td>MEIO-PARA(chave, abrir)</td>
</tr>
<tr>
<td>FINALIDADE-DE(X, Y)</td>
<td>X is a FINALIDADE-DE Y</td>
<td>FINALIDADE-DE(abrir, chave)</td>
</tr>
<tr>
<td>PARTE-DE(X, Y)</td>
<td>X is a PARTE-DE Y</td>
<td>PARTE-DE(roda, carro)</td>
</tr>
<tr>
<td>INCLUI(X, Y)</td>
<td>X INCLUI Y</td>
<td>INCLUI(carro, roda)</td>
</tr>
<tr>
<td>LOCAL-DE(X, Y)</td>
<td>X is a LOCAL-DE Y</td>
<td>LOCAL-DE(restaurante, comer)</td>
</tr>
<tr>
<td>OCORRE-EM(X, Y)</td>
<td>X OCORRE-EM Y</td>
<td>LOCALIZADO-EM(comer, restaurante)</td>
</tr>
</tbody>
</table>
Building PAPEL

Quantitative studies about the patterns used in the definitions.

Using the chart parser PEN with specific grammars to extract relations.

Semantic relations from definitions:

- $Y \text{ - um tipo de } X \rightarrow \text{HIPERONIMO}(X, Y)$
Detailed example

Causation relation:

- relation between an agent (the causer) and a result (the caused)
- CAUSADOR-DE and its inverse RESULTADO-DE.
- Based on verbs: *causar, originar, provocar, produzir, gerar, motivar, suscitar*.
Examples of (simplified) rules for CAUSADOR-DE:

- \{causad|originad|provocad|producid|gerad|motivad|suscitad\}\{o|os|a|as\} FREQ* PREP CAUSADOR
  - concussão (s.f.) - choque violento originado por uma explosão
    - CAUSADOR-DE(explosão, concussão)
- efeito PREP CAUSADOR
  - maximização (s.f) - efeito de maximizar
    - CAUSADOR-DE(maximizar, maximização)
- devido \{a|ao|à|às|aos\} CAUSADOR
  - engasgo (s.m.) - incapacidade de respirar devido a obstrução da garganta
    - CAUSADOR-DE(obstrução, engasgo)
Detailed example

Examples of (simplified) rules for RESULTADO-DE:

- que {causa|origina|provoca|produz|motiva|gera|suscita} RESULTADO
  - diplodoco (s.m.) - bactéria que causa as meningites cerebrospinais
    - RESULTADO-DE(meningites, diplodoco)
- {causar|originar|provocar|produzir|motivar|gerar|suscitar} RESULTADO
  - incomodidade (s.f.) - o que causa mal-estar ou desconforto
    - RESULTADO-DE(mal-estar, incomodidade)
    - RESULTADO-DE(desconforto, incomodidade)
Detailed example

Examples of current problems:

- Specific relations inside the definition
  - estetoscópio (s.m) – instrumento para auscultar a respiração, as batidas do coração e outros *sons produzidos pelo corpo*.
    - CAUSADOR-DE(corpo, estetoscópio)

- Negation of patterns
  - respeitar (v. tr.) – *não causar* dano
    - RESULTADO-DE(dano, respeitar)
## Current precision

<table>
<thead>
<tr>
<th>Relation name</th>
<th>Runs</th>
<th>Hits</th>
<th>Tagged</th>
<th>Tagged correct</th>
<th>Tagged incorrect</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAUSADOR-DE</td>
<td>4</td>
<td>5657</td>
<td>3354</td>
<td>3222</td>
<td>132</td>
<td>0.96</td>
</tr>
<tr>
<td>RESULTADO-DE</td>
<td>4</td>
<td>1693</td>
<td>1070</td>
<td>972</td>
<td>98</td>
<td>0.91</td>
</tr>
<tr>
<td>FINALIDADE-DE</td>
<td>1</td>
<td>1348</td>
<td>192</td>
<td>184</td>
<td>8</td>
<td>0.96</td>
</tr>
<tr>
<td>MEIO-PARA</td>
<td>1</td>
<td>826</td>
<td>286</td>
<td>258</td>
<td>28</td>
<td>0.9</td>
</tr>
<tr>
<td>INCLUI</td>
<td>2</td>
<td>1908</td>
<td>328</td>
<td>320</td>
<td>8</td>
<td>0.98</td>
</tr>
<tr>
<td>PARTE-DE</td>
<td>2</td>
<td>811</td>
<td>202</td>
<td>171</td>
<td>31</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Number of definitions in the whole DLP = 237,246
Further work

- Improvements to PEN
- Using a broad-coverage parser like PALAVRAS (Bick 2000)
- Validation of semantic relations by people.

Construction of the network:
- grouping the words into synsets
- word sense desambiguation
The end...

Questions/suggestions?

Acknowledgement

- This work was done in the scope of the Linguateca, contract nº339/1.3/C/NAC, project jointly funded by the Portuguese Government and the European Union.
References


