**Relation detection between named entities**

Report of a shared task

Cláudia Freitas, Diana Santos, Cristina Mota,
Hugo Oliveira and Paula Carvalho

NAACL-HLT Conference
Semantic Evaluation Workshop
Boulder, CO, June 2009

**Context**

**Linguateca** (www.linguateca.pt) is a distributed network for fostering the computational processing of the Portuguese language

- Organization of evaluation contests for Portuguese (Morfolimpíadas, HAREM and CLEF [GeoCLEF, QA@CLEF, adhoc CLEF, GIKP, LogCLEF, GIKCLEF])
- Creation of free resources that enable sophisticated processing of Portuguese
- Monitoring and cataloging the area

**Acknowledgement**

- Until December 31, 2008, Linguateca and HAREM were funded through the Portuguese government and the European Union with contract number 339/1.3/C/NAIC
- Now this is work is being pursued thanks to a UMIC grant to FCCN Linguateca

**Motivation**

Advance the semantic processing in Portuguese given the fact there were enough actors interested in relation detection

**Task goals**

- Investigate which relations could be found in texts
- Devise a pilot task to compare systems that recognize those relations

**Background**

**HAREM** - Evaluation of named entity recognition in Portuguese texts

**ReRelEM**

- 10 categories
- 43 types
- 21 subtypes

**HAREM Golden Collection**

- 129 docs
- DOCS: 12
- Paragraphs: 108
- Words: 4417
- NE: 573
- Facets: 642
- ABSTRACCO/abstraction: 17
- CITA/concept: 6
- CON/thing: 5

**What is a relevant relation?**

- The relation needs to make sense in the context of the whole document, even if there are no linguistic clues

**Why?**

- Borders between world knowledge and contextual inference are not always clear
- Annotating all possible relations is a way of dealing with different informational needs
When I lived in **Peru**, I attended a horse show and was able to admire breeds I had known only from pictures before, like Falabella and **Paso**.

**relation of breeds to horse shows: are local breeds predominant?**

---

**what is a relevant relation?**

---

**Relation inventory**

- **Identity (ident)**
  - foi fundada em 1131 por D. Telo (São Teotónio)
  - Os adeptos do Porto invadiram a cidade do Porto em júbilo

- **Inclusion (inclui (includes) / incluido (included))**
  - Hamilton, colega de Alonso na McLaren

- **Placement (ocorre-em (occurs-in) / sede-de (place_of))**
  - GP Brasil – Não faltou emoção em Interlagos no Circuito José Carlos Pace desde a primeira volta...

---

**Relation description**

---

**Annotation scheme**

- **Relation inventory**
  - Other (outra)

- **Relation description**
  - Distribution of NE categories for: identity, inclusion, placement.

- **Relation inventory**
  - Distribution of NE categories for: identity, inclusion, placement.

- **Relation description**
  - Other (outra)

---

**Relations and vague categories**

(...) a ideia de uma Europa LOCAL/PESSOA unida. (...) um dia feliz para as cidadãos da **União Europeia** LOCAL. (...) We are mainly a community of values and these common values constitute the foundation of the European Union.
Evaluation

- Evaluate JUST the relations (not the NE)
- Relations with mismatched arguments were ignored
- Alternative segmentations were ignored

Participation and results

<table>
<thead>
<tr>
<th>System</th>
<th>NE task</th>
<th>Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rembrandt</td>
<td>tag</td>
<td>all</td>
</tr>
<tr>
<td>Rembrandt</td>
<td>any identification</td>
<td>all but OUTRA</td>
</tr>
<tr>
<td>SEI Geo</td>
<td>tag</td>
<td>all</td>
</tr>
</tbody>
</table>

- Answer complex questions based on Wikipedia (PhD work in progress)
- Develop a hot news portal based on NEs
- Evaluate a system for ontology creation (final stage of PhD work)

Lessons learned

- Was HAREM grid beneficial to the task of specifying relations? Yes
  Astrônomos brasileiros esperam fotografar os primeiros planetas fora do Sistema Solar com a ajuda do maior telescópio do mundo, o Gemini (...) os telescópios Gemini têm capacidade científica...
  Brazilian astronomers expect to take the first pictures of planets beyond the solar system with the help of the largest telescope in the world, Gemini (...) Gemini telescopes have capacity...

- Relation between COISA-MEMBROCLASSE and COISA-CLASSE is more informative than simply between COISA

- It is easier for people to assign specific relations than too general ones (OUTRA,...)

Food for thought

Beware of too much emphasis on inter-annotator agreement (IAA)

- Is IAA an obstacle to interesting research on semantics by implicitly claiming that anything that is not consensual cannot be there?
- Are resources with IAA better or more useful than resources with innovative semantic information created by merging different human intuitions?
From ReRelEM to ?

LÂMPADA – Second HAREM Resource Package
http://www.linguatexta.pt/HAREM/PacoteRecursos/SegundoHAREM.zip

- Collections
- Evaluation programs
- System runs

Third HAREM
- More realistic user models, possibly within specialized domains (such as cooking, medicine and/or politics)
- Crosslingual IE