

Relation detection between named entities

Report of a shared task

Cláudia Freitas, Diana Santos, Cristina Mota,
Hugo Oliveira and Paula Carvalho
Linguateca / FCCN
Portugal

NAACL-HLT Conference
Semantic Evaluation Workshop
Boulder, CO, June 4th, 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** (Morfolimpiadas, HAREM and CLEF [GeoCLEF, QA@CLEF, adhoc CLEF, GikiP, LogCLEF, GikiCLEF])
- ❖ Creation of free resources that enable sophisticated processing of Portuguese
- ❖ Monitoring and cataloguing 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/NAC
- Now this work is being pursued thanks to a UMIC grant to FCCN

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

Motivation

Anaphora resolution
Mitkov, 2000; Colloveni et al., 2007; de Souza et al. 2008

Focused on co-reference
Anaphoric chains in texts

+

Relation detection
Agichtein and Gravano, 2000; Zhao and Grishman, 2005; Culotta and Sorensen, 2004

Fact extraction
World knowledge

=

ReReIEM
Reconhecimento de Relações entre Entidades Mencionadas
Relation detection between named entities

Background

HAREM - Evaluation of named entity recognition in Portuguese texts

Call for participation: September 2007
Proposal of 3 tasks: November 2007
Release of training material: January 2008
Submission period: April 2008
Workshop: September 2008

- 10 categories
 - 43 types
 - 21 subtypes

HAREM Golden Collection
129 docs

ReReIEM Golden Collection

Category/gloss	#
PESSOA/person	196
LOCAL/place	145
OR	102
TEX	84
DOCS: 12	33
Paragraphs: 108	33
OB	33
Words: 4417	21
VA	17
NE: 573	6
ACI	5
Facets: 642	
ABSTRACCAO/abstraction	
OUTRO/other	
COISA/thing	

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

What is a relevant relation?

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?

Relation inventory

Identity (ident)

✓ foi fundada em 1131 por **D. Telo (São Teotónio)**
It was founded in 1132 by D. Telo (São Teotónio)

✗ Os adeptos do **Porto** invadiram a cidade do **Porto** em júbio
The (FC) Porto fans invaded the (city of) Porto, very happy

Inclusion (inclui (includes) / incluído (included))

Hamilton, colega de Alonso na **McLaren**
Lewis Hamilton, Alonso's team-mate in 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...

GP Brasil – There was no lack of excitement in Interlagos at the José Carlos Pace Circuit

Relation inventory

Other (outra)

Relation / gloss	#
vinculo-inst / inst-commitment	936
obra-de / work-of	300
participante-em / participant-in	202
relacao-participacao-de / has-participant	202
relacao-familiar / family-tie	90
residencia-de / home-of	75
natural-de / born-in	47
relacao-profissional / professional-tie	46
povo-de / people-of	30
representante-de / representative-of	19
residente-de / living-in	18
personagem-de / character-of	12
periodo-vida / life-period	11
propriedade-de / owned-by	10
proprietario-de / owner-of	10
representado-por / represented-by	7
estabelecido-em / established-in	7
outra-rel / other	6
nome-de-ident / name-of	4
outra-edicao / other-edition	2

Annotation scheme

Diagram illustrating an annotation scheme with examples of relations and their types. Relations are shown as red boxes with arrows indicating the relation type (e.g., 'outra', 'inclui', 'ident').

```

<EM ID="ex1-39" CATEG="PESSOA" TIPO="INDIVIDUAL" >Miguel Rodrigues</EM>,
chefe dos
<EM ID="ex1-40" CATEG="ORGANIZACAO" TIPO="INSTITUICAO" COREL="ex1-39"
TIPOREL="outra" >Serviços Administrativos</EM>
da
<EM ID="ex1-41" CATEG="ORGANIZACAO" TIPO="INSTITUICAO" COREL="ex1-40"
TIPOREL="inclui" >Universidade de Trás-os-Montes e Alto Douro</EM>.
<EM ID="ex1-42" CATEG="ORGANIZACAO" TIPO="INSTITUICAO" COREL="ex1-41"
TIPOREL="ident" >UTAD</EM>.
    
```

No need to annotate all relations

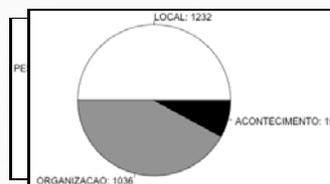
Evaluation program expands all possible relations

Vilain et al. (1995)

A ident B ^ B ident C	⇒ A ident C
A inclui B ^ B inclui C	⇒ A inclui C
A inclui B ^ B sede_de C	⇒ A sede_de C
A ident B ^ B any_rel C	⇒ A any_rel C

Relation description

Relations: 6790
Identity: 1436
Inclusion: 1612
Placement: 1232
Other: 2510



Distribution of NE categories for **Inclusion**

Relations and vague categories

(...) a ideia de uma **Europa** **unida**. (...) um dia feliz para as cidadãs e os cidadãos da **União Europeia**. (...) Somos essencialmente **uma** comunidade de valores -- são estes valores comuns que **constituem** o fundamento da **União Europeia**.

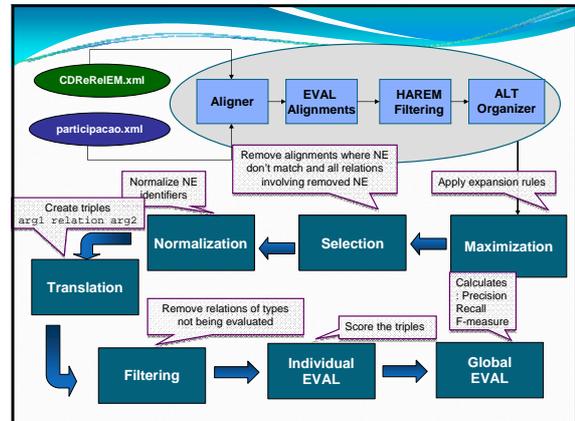
the idea of a united Europe (they) a happy day for the citizens of the European Union (...) We are mainly a community of values' and these common values constitute the foundation of the European Union

Evaluation

- Evaluate JUST the relations (not the NE)

Relations with mismatched arguments were ignored

Alternative segmentations were ignored



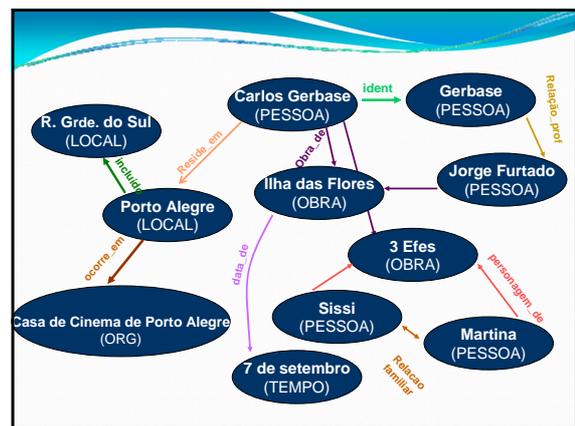
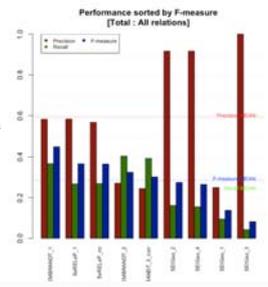
Participation and results

System	NE task	Relations
Rembrandt	all	all
SeRelEP	only identification	all but out:za
SaiGeo	only LOCAL detection	inclusion

- 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? ✓

Astrónomos 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 a 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 ReReIEM to ?



LÂMPADA – Second HAREM Resource Package

<http://www.linguateca.pt/HAREM/PacoteRecursosSegundoHAREM.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