Relation detection between named entities: report of a shared task

Cláudia Freitas, Diana Santos
Cristina Mota
SINTEF ICT
claudiafreitas@puc-rio.br
Diana.Santos@sintef.no
cmota@ist.utl.pt

Hugo Gonçalo Oliveira

CISUC, DEI - FCTUC
hroliv@dei.uc.pt

Paula Carvalho

Univ. Lisbon, FCUL, XLDB
pcc@di.fc.ul.pt

Abstract

In this paper we describe the first evaluation contest (track) for Portuguese whose goal was to detect and classify relations between named entities in running text, called ReRelEM. Given a collection annotated with named entities belonging to ten different semantic categories, we marked all relationships between them within each document. We used the following fourfold relationship classification: identity, included-in, located-in, and other (which was later on explicitly detailed into twenty different relations). We provide a quantitative description of this evaluation resource, as well as describe the evaluation architecture and summarize the results of the participating systems in the track.

1 Motivation

Named entity recognition can be considered the first step towards semantic analysis of texts and a crucial subtask of information extraction systems. Proper names, besides their high frequency in language, do more than just refer – they convey additional information as instances of general semantic categories. But NE recognition is, as just mentioned, only the first step for full language processing. If we want to go beyond the detection of entities, a natural step is establishing semantic relations between these entities, and this is what this paper is about.

There are two fairly independent communities that focus on the task of detecting relations between named entities: the work on anaphora resolution, illustrated by (Mitkov, 2000; Collovini et al., 2007;

de Souza et al., 2008) and the work on relation detection in information extraction, see e.g. (Agichtein and Gravano, 2000; Zhao and Grishman, 2005; Culotta and Sorensen, 2004). Although both communities are doing computational semantics, the two fields are largely non-overlapping, and one of the merits of our work is that we tried to merge the two.

Let us briefly describe both traditions: as (Mitkov, 2000) explains, anaphora resolution is concerned with studying the linguistic phenomenon of pointing back to another expression in the text. The semantic relations between the referents of these expressions can be of different types, being co-reference a special case when the relation is identity. The focus of anaphora resolution is determining the antecedent chains, although it implicitly also allows to elicit semantic relations between referents. This task has a long tradition in natural language processing (NLP) since the early days of artificial intelligence (Webber, 1978), and has from the start been considered a key ingredient in text understanding.

A different tradition, within information extraction and ontology building, is devoted to fact extraction. The detection of relations involving named entities is seen as a step towards a more structured model of the meaning of a text. The main concerns here (see e.g. (Zhao and Grishman, 2005)) are the extraction of large quantities of facts, generally coupled with machine learning approaches.[1]

Although mentions of named entities may ex-

---

[1]Other authors use the term *relation detection* in still other ways: for example, (Roth and tau Yih, 2004) use it for the translation of any natural language sentences into "logical form", as in *kill (x,y)*. This task does not concern us here.

Relations	Works
orgBased-in, Headquarters, Org-Location, Based-in	RY, AG, DI, Sn, CS, ACE07, ACE04, ZG
live-in, Citizen-or-Resident	RY, ACE04, ZG, ACE07,CS
Employment, Membership, Subsidiary	ZG, CS, ACE04, ACE07
located(in), residence, near	ACE04, ACE07,CS, ZG
work-for, Affiliate, Founder, Management, Client, Member, Staff	CS, ACE04, ACE07, RY, ZG
Associate, Grandparent, Parent, Sibling, Spouse, Other-professional, Other-relative, Other-personal	CS, ACE04, ACE07
User, Owner,Inventor, Manufacturer	ACE04,ACE07, ZG, CS
DiseaseOutbreaks	AG
Metonymy	ACE07
identity	ARE
synonym	ARE
generalisation	ARE
specialisation	ARE

Table 1: Relations used in other works or evaluation contests.

press semantic relations other than identity or dependency, the main focus of the first school has been limited to co-reference. Yet, relations such as `part-of` have been considered under the label of indirect anaphora, also known as associative or bridging anaphora.

Contrarywise, the list of relations of interest for the second school is defined simply by world knowledge (not linguistic clues), and typical are the relations between an event and its location, or an organization and its headquarters. Obviously, these relations do occur between entities that do not involve (direct or indirect) anaphora in whatever broad understanding of the term.

Also, relation detection in the second school does not usually cover identity (cf. ACE's seven relation types): identity or co-reference is often considered an intermediate step before relation extraction (Culotta and Sorensen, 2004).

Table 1 displays a non-exhaustive overview of the different relations found in the literature.[2]

In devising the ReRelEM[3] pilot track, our goal was twofold: to investigate which relations could be found between named entities in Portuguese text, and how could a pilot task be devised that compared the performance of different automatic systems supposed to identify them. It should be emphasized that both MUC and ACE were key inspiration sources for ReRelEM, which stems from Linguateca's emphasis on evaluation.

In fact, we were conversant with MUC co-reference track and the way it was scored, as well as aware of two other related evaluation contests: ACE (Doddington et al., 2004; NIST and ACE, 2007), which extended MUC by dropping the requirement that entities had to be named, and ARE (Orăsan et al., 2008), which requested the identification of an anaphoric relation in certain types of pre-defined relations (identity, synonymy, generalization and specification), but which ignored indirect anaphora (that may convey meronymy, or inclusion, in a broad sense).

ReRelEM, although maintaining (or adding) the restriction to named entitites, is, from our point of view, an advance in the field of relation detection, since we proposed the detection (and classification) of all (relevant) kinds of relations between NEs in a document, providing thus both a merge and an extension of the previous evaluation campaigns.

---

[2]There is overlap between ACE 2007 and 2004 types of relations. In order to ease the comparison, we used the names of subtypes for ACE relations.

[3]ReRelEM stands for *Reconhecimento de Relações entre Entidades Mencionadas*, Portuguese for "recognition of relations between named entities", see (Freitas et al., 2008).

Category/gloss	#
PESSOA/person	196
LOCAL/place	145
ORGANIZACAO/org	102
TEMPO/time	84
OBRA/title	33
VALOR/value	33
ACONTECIMENTO/event	21
ABSTRACCAO/abstraction	17
OUTRO/other	6
COISA/thing	5

Table 2: Category distribution in the golden collection

2 Track description

The purpose of ReRelEM is to assess systems that try to recognize the most relevant relations between named entities, even if those relations do not involve coreference or anaphora.

2.1 Context

In order for it to be feasible in the short time we had, the track definition required that both referring expression and their semantic referent were named entities. Pronouns and definite descriptions were hence excluded. Note also that ReRelEM was defined in the context of the second edition of a larger evaluation contest dealing with NE detection and classification in Portuguese, HAREM (Santos et al., 2008) (for a detailed description of HAREM, in Portuguese, see also (Santos and Cardoso, 2007; Mota and Santos, 2008)). HAREM required systems to choose among ten categories (see Table 2), 43 types and 21 subtypes, the later concerning the categories TEMPO (time) and LOCAL (place).

So, it should be emphasized that ReRelEM focuses only on the classification and detection of the relations, not limiting in any way the kinds of (named) entities that can be related (as usualy done in other detection tasks). It only enforces the kinds of relations that must be identified.

2.2 Relation inventory

The establishment of an inventory of the most relevant relations between NEs is ultimately subjective, depending on the kind of information that each participant aims to extract. We have nevertheless done an exploratory study and annotated exhaustively a few texts to assess the most frequent and less controversial (or easier to assign) relations, and came up with just the following relation types for the task proposal:

- identity (`ident`);

- inclusion (`inclui` (includes) or `incluido` (included));

- placement (`ocorre-em` (occurs-in) or `sede-de` (place-of));

- other (`outra`)

For further description and examples see section 3.

However, during the process of building ReRelEM's golden collection (a subset of the HAREM collection used as gold standard), human annotation was felt to be more reliable – and also more understandable – if one specified what "other" actually meant, and so a further level of detail (twenty new relations) was selected and marked, see Table 3. (In any case, since this new refinement did not belong to the initial task description, all were mapped back to the coarser `outra` relation for evaluation purposes.)

2.3 ReRelEM features and HAREM requirements

The annotation process began after the annotation of HAREM's golden collection, that is, the relations started to be annotated after all NE had been tagged and totally revised. For ReRelEM, we had therefore no say in that process – again, ReRelEM was only concerned with the relations between the classified NEs. However, our detailed consideration of relations helped to uncover – and correct some mistakes in the original classification.

In order to explain the task(s) at hand, let us describe shortly ReRelEM's syntax: In ReRelEM's golden collection, each NE has a unique ID. A relation between NE is indicated by the additional attributes COREL (filled with the ID of the related entity) and TIPOREL (filled with the name of the relation) present in the NE that corresponds to one of the arguments of the relation. (Actually, there's no difference if the relation is marked in the first or in the second argument.)

131

One referring expression can be associated with one or more NEs through several semantic relations. In such cases, all possible relations must be assigned to the referring expression, in the form of a list, as illustrated by *UTAD* in Figure 1.

In this example, the NE with name *UTAD* (and id `ex1-42`) corresponds to an acronym of *Universidade de Trás-os-Montes e Alto Douro* (a university in Portugal), maintaining with this entity an identity relation (`ident`). The TIPOREL field of `ex1-42` contains another relation, `inclui`, which stands for the relation of inclusion, this time with the previously mentioned *Serviços Administrativos* (`ex1-40`), a specific department of the university.

In order to minimize human labour and also to let systems mark relations the way it would better suit them, we have postulated from the start that, for all purposes, it would be equivalent to annotate everything or just enough relations so that all others can be automatically computed. So, the evaluation programs, in an obvious extension of what was proposed in (Vilain et al., 1995) for identity,

1. add/expand all relations with their inverses (e.g., "A includes B" entails "B is included in A"), and

2. apply a set of expansion rules (see examples in Table 4) to compute the closure

As a consequence, different systems may tag the same text in different ways, but encoding the same knowledge.

2.4 What is a relevant relation?

An important difference as to what we expect as relevant relations should be pointed out: instead of requiring explicit (linguistic) clues, as in traditional research on anaphor, we look for all relations that may make sense in the specific context of the whole document. Let us provide two arguments supporting this decision:

- the first one is philosophical: the borders between world knowledge and contextual inference can be unclear in many cases, so it is not easy to distinguish them, even if we did believe in that separation in the first place;

- the second is practical: marking all possible relations is a way to also deal with unpredictable informational needs, for example for text mining applications. Take a sentence like "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.". From this sentence, few people would infer that Paso is a Peruvian breed, but a horse specialist might at once see the connection. The question is: should one identify a relation between Peru and Paso in this document? We took the affirmative decision, assuming the existence of users interested in the topic: "relation of breeds to horse shows: are local breeds predominant?".

However, and since this was the first time such evaluation contest was run, we took the following measure: we added the attribute INDEP to the cases where the relation was not possible to be inferred by the text. In this way, it is possible to assess the frequency of these cases in the texts, and one may even filter them out before scoring the system runs to check their weight in the ranking of the systems. Interestingly, there were very few cases (only 6) marked INDEP in the annotated collection.

3 Qualitative relation description

Identity (or co-reference) does not need to be explained, although we should insist that this is not identity of expression, but of meaning. So the same string does not necessarily imply identity, cf.:

> Os adeptos do **Porto** invadiram a cidade do **Porto** em júbilo.[4]

Interestingly, even though organization is only the third most frequent category, Figure 2 shows that we found more co-reference among organizations than among any other category.

As to inclusion (see Figure 3), it was defined between NEs of the same sort, as the following examples, respectively illustrating LOCAL, PESSOA, OBRA and ORGANIZACAO, show:

> Centenas de pessoas recepcionaram no **Aeroporto da Portela** num clima de enorme entusiasmo e euforia, a selecção

---

[4]The (FC) Porto fans invaded the (city of) **Porto**, very happy

```
<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 ex1-40"
TIPOREL="ident inclui">UTAD</EM>
```
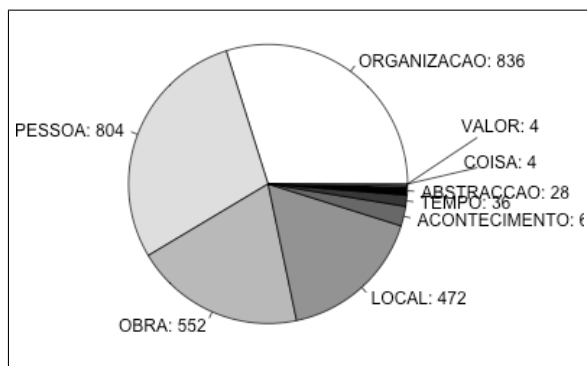
Figure 1: Full example of ReRelEM syntax.
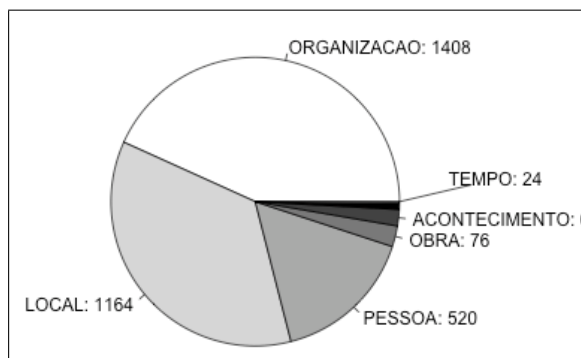

Figure 2: Distribution of NE categories for identity.

Figure 3: NE categories related by inclusion.

portuguesa de râguebi. A boa prestação global da equipa (...) não passou despercebida em **Portugal**.[5]

Lewis Hamilton, colega de Alonso na **McLaren**[6]

da assinatura do **Tratado de Lisboa** (...) de ver reconhecido o valor juridicamente vinculativo da **Carta** um passo "essencial no quadro de reforma dos **Tratados**[7]

por participar na cerimónia de proclamação da Carta dos Direitos Fundamentais da **UE** (...) salientou ainda o compromisso assumido pelas três instituições - **PE**[8]

---

[5]Hundreds of people waited with enthusiasm and euphoria at the **Portela Airport** for the Portuguese national rugby team.(...) The team's good performance did not go unnoticed in **Portugal**

[6]**Lewis Hamilton**, Alonso's team-mate in **McLaren** – Note that, in HAREM, teams are considered groups of people, therefore an individual and a team have the same category PESSOA (person), but differ in the type.

[7]the signing of the **Lisbon Treaty** (...) juridically vinculative value of the **Charter**, a crucial step for the **Treaties** reform policy

[8]to participate in the proclamation ceremony of the Charter

Placement is clearly skewed towards placement of organizations (518 cases) as opposed to occurrence of events (just 98 instances). However, if we consider the relative distribution of organizations and events (see Table 2), we can state that, relative to their places, events have 4.8 relations in average and organizations 5.0, which is a far more interesting result, not favouring any of the NE classes.

Examples of this relation are:

GP Brasil – Não faltou emoção em Interlagos no **Circuito José Carlos Pace**[9]

As to the refinement of `outra`, Table 3 presents the relations found in the material.

3.1 Vague categories

It is important to stress that the basic tenets of HAREM had to be followed or reckoned with, not only the classification grid (see Table 2) but particularly the fact that some named entities are considered to be vague among different categories in

---

of Fundamental Rights of the **EU** (...) stressed the commitment assumed by the three institutions - **EP**

[9]**GP Brasil** – There was no lack of excitement in Interlagos at the **José Carlos Pace Circuit**.

Relation / gloss	Number
vinculo-inst / inst-commitment	936
obra-de / work-of	300
participante-em / participant-in	202
ter-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	15
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
praticado-em / practised-in	7
outra-rel/other	6
nome-de-ident / name-of	4
outra-edicao / other-edition	2

Table 3: Frequency of `other` relations.

HAREM.[10]

This last property, namely that named entities could belong to more than one category, posed some problems, since it was not straightforward whether different relations would involve all or just some (or one) category. So, in order to specify clearly the relations between vague NEs, we decided to specify separate relations between facets of vague named entities. Cf. the following example, in which vagueness is conveyed by a slash:

> (...) a ideia de uma **Europa** (LO-CAL/PESSOA) unida. (...) um dia feliz para as cidadãs e os cidadãos da **União Europeia** (LOCAL). (...) Somos essencialmente uma comunidade de valores – são estes valores comuns que constituem o fundamento da **União Europeia** (AB-STRACCAO/ORG/LOCAL).[11]

The several relations between the three bold-faced NEs have been found to be as follows: The LOCAL facet of the first NE is identical with the LOCAL facets of the second and third NEs, while the ORG(ANIZACAO) facet of the third NE is located in the LOCAL facet of the second and first NEs. (Two kinds of relations are therefore involved here: `ident` and `inclui`.)

4 Evaluation: architecture and measures

Our first concern in this pilot track was to make a clear separation between the evaluation of relations and the evaluation of NE detection, which was the goal of HAREM. So, ReRelEM 's evaluation uses as a starting point the set of alignments that correspond to a mapping of the NE in the golden collection (GC) to a (candidate) NE in the participation.

Evaluation has the following stages:

- Maximization: the sets of relations annotated in both the GC and in the participation are maximized, applying the rules in Table 4;

- Selection: the alignments where the NE in the GC is different from the corresponding one in the participation are removed, and so are all relations held between removed NEs;

- Normalization: The identifiers of the NE in the participation are normalized in order to make it possible to compare the relations in both sides, given that each system uses its own identifiers.

- Translation: The alignments are translated to triples: `arg1 relation arg2`, where the arguments consist of the identifiers of the NE together with the facet, for example `x67 LOCAL sede-de ty45 ORGANIZACAO`.

- Filtering: removing relations of types not being evaluated (because HAREM, and therefore ReRelEM, allows for partial participation – and evaluation – scenarios[12]).

- Individual evaluation: the triples in the GC are compared to the triples in the participation.

---

[10] This is different from considering metonymy classes, in that no classifications are considered more basic than others, see (Santos, 2006) for vagueness as an intrinsic property of natural language.

[11] the idea of a united **Europe** (...) 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**.

[12] In other words, it is possible to select a subset of the classification hierarchy.

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		

Table 4: Maximization rules

System	NE task	Relations
Rembr.	all	all
SeRelEP	only identification	all but outra
SeiGeo	only LOCAL detection	inclusion

Table 5: Participant systems

- Global evaluation: measures (precision, recall and F-measure) are calculated based on the score of each triple.

Each triple is scored as correct, missing or incorrect. We only considered as correct triples (and correct relations) those which linked the correct NEs and whose relation was well classified. So, a system doesn't score if it correctly matches the NEs to be related, but fails to recognize the kind of relation. We assign one point to each correct relation and none to incorrect or missing relations, and then we compute precision, recall and F-measure.

ReRelEM's golden collection includes 12 texts with 4,417 words and 573 NEs (corresponding to 642 different facets). In all we annotated 6,790 relations (1436 identity; 1612 inclusion; 1232 placement; 2510 other).

5 Participation and results

For this first edition, only three systems (totalling nine runs) participated, namely REMBRANDT (Cardoso, 2008), SEI-Geo (Chaves, 2008), and SeRelEP (Bruckschen et al., 2008), whose results are found in Figure 4. However, they did not compare well: they selected different NER tasks and different relation types, as shown in Table 5. So, given the little and diverse participation in ReRelEM, we cannot do a useful state of the art, but we were definitely able to provide an interesting and important resource for empirical studies and for training of future systems, as well as a set of publicly available programs to manipulate, evaluate and display this
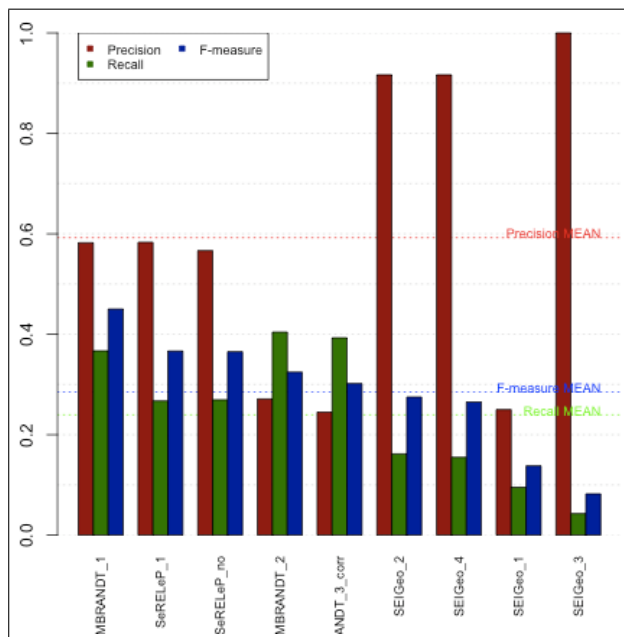

Figure 4: ReRelEM results: F-measure, all relations

kind of semantic data[13].

6 Discussion and further work

Although this was just a pilot, a lot of knowledge about the task and the problems to be dealt with were gathered for the future, and important resources were offered to the community.

We intend to annotate further sections of the HAREM golden collection (as well as other kinds of texts and materials) with more relations in order to have more quantitative empirical data for studying the semantic fabric of Portuguese.

Although from an organization point of view it made sense to couple ReRelEM with HAREM, one should reflect over the consequences of inheriting a lot of decisions taken in HAREM, somehow going counter the intuitive and easier task of just annotating relations in a first round. However, despite initial fears to the contrary, we found out that the considerably fine-grained HAREM grid was in fact beneficial to the task of specifying relations: it is, after all, much more informative to have a relation of inclusion between a COISA-MEMBROCLASSE (concrete instance of a class of objects) and a COISA-CLASSE (a class of objects), than just a relation of inclusion

---

[13]http://www.linguateca.pt/HAREM/

135

tout court. In fact, in the next sentence, a kind of specialization relation can be uncovered.

> Astrônomos brasileiros esperam fotografar os primeiros planetas fora do Sistema Solar com a ajuda do maior telescópio do mundo, o *Gemini* (...) os telescópios <u>Gemini</u> têm capacidade científica...[14]

Likewise, an inclusion relation held between PESSOA-GRUPOCARGO (a group of roles performed by people) and PESSOA-INDIVIDUAL (an individual person), as in the following example, is more informative than a simple relation of inclusion between NEs, or even inclusion between PESSOA entities without further discrimination.

> *Pöttering*, *Sócrates* e *Barroso* assinam Carta dos Direitos Fundamentais da UE. Depois de a Carta ser assinada pelos <u>Presidentes</u> das três instituições, ouviu-se o hino europeu...[15]

Furthermore, this relation is also different from an inclusion relation held between PESSOA-INDIVIDUAL (an individual) and PESSOA-GRUPOMEMBRO (a group of people):

> *Lobos* recebidos em apoteose. (...) o capitão <u>Vasco Uva</u> explicou por que houve uma empatia tão grande entre... [16]

Conversely, the specification of relations between different NEs in a text may help in detecting and justifying different facets of a particular NE, i.e., multiple semantic categories that should be assigned to it.

This illustrates the often observed case that it may be easier for a human annotator to decide and choose a specific issue than a too general one, and that therefore categories or choices should be more dependent on ease of human interpretation than quantitative factors (such as few categories or balanced ones).

---

[14]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...

[15]**Pöttering**, **Sócrates** e **Barroso** sign the declaration... After being signed by the **Presidents** of the three institutions, ...

[16]**Lobos** received apoteothically. (...) Captain **Vasco Uva** explained why ...

For future work, we obviously intend to increase the size of the annotated collection (to the whole HAREM collection and even beyond), and investigate a couple of issues that interest us: which strategies are used to avoid repetition of proper names and establish textual cohesion? How do relations between noun phrases in general compare with relations between entities?

We would also like to investigate closer relationships between different relations: for example, is it more appropriate to also develop a hierarchy of relations, reconsidering, for example, `affiliation` (currently one of the `other`) as a kind of `inclusion`?

In order to understand better what this task is about, we would also like to investigate whether there are interesting correlations between NE categories and relations, as well as text genre and this sort of connectivity. Even though we only studied and annotated in depth 12 different texts, it was at once obvious that they had quite different properties as far as the number and kinds of relations was concerned.

From an evaluation point of view, we would like to improve our inconsistency detection programs and be able to reason about possible contradictions (of the annotation or of the interpretation) as well as experiment with different weights and evaluation measures, taking into account criteria such as predictability of relationships between NEs.

In any case, we believe this was an important first step to understand a number of issues and to reflect about what computational systems should be doing to harvest semantic knowledge. We would like to receive feedback on whether the task design seems sound to the rest of the community, and whether systems which would perform well in such task could be put to good use in real world applications.

Acknowledgments

References

Eugene Agichtein and Luis Gravano. 2000. Snowball: Extracting Relations from Large Plain-Text Collections. In *Proc. of the 5th ACM International Conference on Digital Libraries (ACM DL)*, pages 85–94, San Antonio, Texas, USA, June, 2-7.

Mírian Bruckschen, José Guilherme Camargo de Souza, Renata Vieira, and Sandro Rigo. 2008. Sistema SeRELeP para o reconhecimento de relações entre entidades mencionadas. In Mota and Santos (Mota and Santos, 2008).

Nuno Cardoso. 2008. REMBRANDT - Reconhecimento de Entidades Mencionadas Baseado em Relações e ANálise Detalhada do Texto. In Mota and Santos (Mota and Santos, 2008).

Marcírio Chaves. 2008. Geo-ontologias e padrões para reconhecimento de locais e de suas relações em textos: o SEI-Geo no Segundo HAREM. In Mota and Santos (Mota and Santos, 2008).

Sandra Collovini, Thiago Ianez Carbonel, Juliana Thiesen Fuchs, Jorge César Coelho, Lucia Helena Machado Rino, and Renata Vieira. 2007. Summ-it: Um corpus anotado com informações discursivas visando à sumarização automática. In *Anais do XXVII Congresso da SBC: V Workshop em Tecnologia da Informação e da Linguagem Humana – TIL*, pages 1605–1614, Rio de Janeiro, RJ, Brazil, junho/julho.

Aron Culotta and Jeffrey Sorensen. 2004. Dependency tree kernels for relation extraction. In *Proceedings of the 42rd Annual Meeting of the Association for Computational Linguistics (ACL'04)*, pages 423–429. Association for Computational Linguistics, July.

José Guilherme Camargo de Souza, Patrícia Nunes Gonçalves, and Renata Vieira. 2008. Learning coreference resolution for portuguese texts. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira, and Paulo Quaresma, editors, *PROPOR*, volume 5190 of *Lecture Notes in Computer Science*, pages 153–162. Springer.

Georde Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ace) programm. tasks, data and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, pages 837–840, Lisbon, Portugal.

Cláudia Freitas, Diana Santos, Hugo Gonçalo Oliveira, Paula Carvalho, and Cristina Mota. 2008. Relações semânticas do ReRelEM: além das entidades no Segundo HAREM. In Mota and Santos (Mota and Santos, 2008).

Ruslan Mitkov. 2000. Towards a more consistent and comprehensive evaluation of anaphora resolution algorithms and systems. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC-2000)*, pages 96–107, Lancaster, UK.

Cristina Mota and Diana Santos, editors. 2008. *Desafios no reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguateca.

NIST and ACE. 2007. Automatic Content Extraction 2008 Evaluation Plan (ACE08) – Assessment of Detection and Recognition of Entities and Relations within and across Documents. Technical report, NIST.

Constantin Orăsan, Dan Cristea, Ruslan Mitkov, and Antonio Branco. 2008. Anaphora resolution exercise: An overview. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May, 28 - 30.

Dan Roth and Wen tau Yih. 2004. A linear programming formulation for global inference in natural language tasks. In *Proceedings of CoNLL-2004*, pages 1–8.

Diana Santos and Nuno Cardoso, editors. 2007. *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*. Linguateca, Portugal.

Diana Santos, Cláudia Freitas, Hugo Gonçalo Oliveira, and Paula Carvalho. 2008. Second HAREM: new challenges and old wisdom. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira, and Paulo Quaresma, editors, *Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008)*, volume LNAI 5190, pages 212–215. Springer Verlag.

Diana Santos. 2006. What is natural language? Differences compared to artificial languages, and consequences for natural language processing, 15 May. Invited lecture, SBLP2006 and PROPOR'2006.

Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A model-theoretic coreference scoring scheme. In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 45–52. Morgan Kaufmann.

Bonnie Lynn Webber. 1978. *A formal approach to discourse anaphora*. Outstanding dissertations in linguistics. Garland Publishing, New York, NY, USA.

Shubin Zhao and Ralph Grishman. 2005. Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL 2005)*, pages 419–426, Morristown, NJ, USA. Association for Computational Linguistics.