

# Where Anaphora and Coreference Meet. Annotation in the Spanish CESS-ECE Corpus

Marta Recasens, M. Antònia Martí, Mariona Taulé  
CLiC, Centre de Llenguatge i Computació  
University of Barcelona  
Gran Via de les Corts Catalanes, 585  
Barcelona 08007, Spain  
{mrecasens,amarti,mtaule}@ub.edu

## Abstract

This paper describes the guidelines of the annotation scheme designed to enrich the Spanish CESS-ECE corpus with coreference information, which is a significant step towards the definition of an exhaustive typology of pronominal and full NP coreferential expressions and their relations for Spanish. The goal is twofold. From a computational perspective, this work establishes the formal foundations for the construction of the largest corpus of Spanish texts annotated from the morphological to the pragmatic level. This corpus, which will be publicly released, will be used to construct an automatic corpus-based coreference resolution system. From a linguistic point of view, hypotheses on coreferential expressions will be tested and validated on this framework.

## Keywords

Coreference resolution, anaphora resolution, corpus linguistics, annotation scheme.

## 1 Introduction

Natural Language Processing (NLP) applications such as information extraction, text summarization and question answering need to identify all the information that is said about one same entity throughout a text. Consequently, systems capable of resolving coreference –and, by extension, anaphora– are essential. There are basically two approaches: knowledge-based and corpus-based. However, as pointed out by Mitkov [11], “corpora annotated with anaphoric or coreferential links are still a rare commodity, and those that do exist are not of a large size.” Specifically, in Spanish, the field of computational coreference resolution is still highly knowledge-based.

With a view to building a corpus-based coreference resolution system for Spanish, our project is to extend the morphologically, syntactically and semantically annotated CESS-ECE corpus (500,000 words) with pronominal and full NP coreference information. We believe that the more consistent the linguistic basis underlying the annotation scheme is, the easier it is to build a state-of-the-art coreference resolution system. On the other hand, coreferential –anaphoric in particular– relations are very much specific to each language. Unlike English, for instance, Spanish has zero and clitic pronouns. Therefore, it is fundamental to define the typology of expressions (pronouns, full NPs and proper nouns) that can enter in coreferential relations in Spanish as well as the types of

relations.<sup>1</sup> This typology forms the basis for a flexible markup scheme, rich enough to cover the cases of coreference in Spanish.

Apart from being a useful resource for training and evaluating coreference resolution systems for Spanish; from a linguistic point of view, the annotated corpus will serve as a workbench to test for Spanish the hypotheses suggested by Ariel [1] and Gundel et al. [6] about the cognitive factors governing the use of referring expressions. The only way theoretical claims coming from a single person’s intuitions can be proved is on the basis of empirical data that have been annotated in a reliable way.

This paper lays the foundations for our ongoing project. Taking the CESS-ECE corpus as the starting point, we describe the adaptation of the MATE meta-scheme for anaphora annotation [15] by considering both the information already codified in the CESS-ECE corpus and the way new information should be annotated. Sound linguistic criteria guide the decisions made throughout the process.

The rest of the paper proceeds as follows: Section 2 delimits the frame of the coreferential and anaphoric phenomena that we deal with. A brief state of the art of anaphora resolution systems existing for Spanish is provided in Section 3. The guidelines of our annotation scheme and methodology are given in Section 4. Finally, Section 5 presents our conclusions and further work.

## 2 Coreference along a continuum

Anaphora is the linguistic phenomenon by which a word is interpreted with the help of some previous item (the antecedent) in the discourse. The anaphor and the antecedent may be coreferential or “colexical” –coinage of our own–, that is, they may have the same discourse referent (1a) or just share the semantic type (1b).

- (1) a. Llegaron con buenos resultados hasta los torneos de la final, pero en ellos perdieron.
- b. Los mejores equipos de la NBA son mejores que los nuestros.
- c. La capital de Francia...en París...<sup>2</sup>

---

<sup>1</sup> Anaphora in Spanish has been mainly studied from a descriptive grammar point of view [3]. From a pragmatic perspective, the recent study by Blackwell [2] tests the neo-Gricean maxims on the basis of oral data.

Coreference and anaphora are thus closely interrelated, although not all anaphoric relations are coreferential (1b), nor are all coreferential relations anaphoric (1c). Our main concern is coreference; as regards colexicality then, we limit ourselves to solving those colexical anaphors occurring in headless definite NPs in order to recover the semantic type of the head, as they may be part of a coreferential chain.

When speakers solve a coreferential link, they rely –to a greater or lesser extent– on linguistic or/and world knowledge. The more anaphoric a coreferential expression is, the more linguistic knowledge is required; the less anaphoric, the more world knowledge. The expression of coreference is best seen as a continuum, ranging from zero and anaphoric pronouns to self-sufficient definite descriptions (DD, see Section 4.2.1) and proper nouns (Figure 1). Fully aware that it is a too coarse simplification, we propose this gradation just as a starting point. One of the goals of our project from a linguistic point of view is to achieve a better understanding of the grades of this continuum.

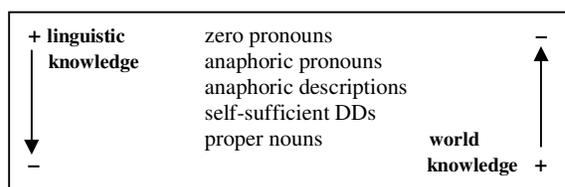


Figure 1: Range of expressions creating coreference

The piece of news in (2) illustrates different expressions referring to one same entity: *the Canary Islands*.

(2) El número de parados registrado en Canarias en mayo subió y la cifra de desempleados en las islas se sitúa hoy en 89.764. Aunque es todavía pronto para sacar conclusiones, los políticos de la comunidad canaria ya han apuntado posibles causas y no descartan giros inesperados en la economía de la zona. De hecho, Ø es un territorio del que los periódicos suelen hablar, pero no precisamente de su tasa de paro.<sup>3</sup>

The entity is first evoked in the discourse by means of a proper noun (*Canarias*) and it is next expressed via an anaphoric description (*las islas*) –it is the previous proper noun that completes its referential meaning. Later it takes the form of a self-sufficient DD (*la comunidad canaria*), then again an anaphoric description (*la zona*). The last

<sup>2</sup> All translations throughout the paper are literal so as to make the Spanish wording as transparent as possible.

- (1) a. They got with good results to the final competitions, but they lost in them.
- b. The best teams of the NBA are better than ours.
- c. The capital of France...in Paris...

<sup>3</sup> (2) The number of unemployed people recorded in the Canary Islands in May increased and the number of unemployed in the islands is today 89,764. Although it is still early to draw conclusions, the politicians of the Canarian Community have already suggested possible causes and do not discard unexpected turns in the economy of the area. In fact, (it) is a region about which newspapers usually talk, but not precisely about its unemployment rate.

three elements of the coreference chain are a zero subject pronoun, a relative pronoun (*que*) and a possessive (*su*).

The whole of this continuum is the basis for the typology of coreferential expressions that our project focuses on.

### 3 Coreference resolution in Spanish

The computational coreference resolution in Spanish has been restricted to the resolution of third person anaphoric and zero pronouns [14] and to the resolution of descriptions introduced by the definite article or a demonstrative that corefer with another NP [13] by applying heuristics on shallowly parsed texts. Evaluated on a corpus containing 1,217 descriptions, Muñoz's [13] algorithm achieved 79.5% precision. Saiz-Noeda's [18] ERA system is an extension of the algorithm of [14], which, in turn, is an adaptation to Spanish of the set of constraints and preferences used by Lappin & Leass [9] in their system for English. Palomar et al.'s [14] algorithm makes use of lexical, morphological and syntactic information (partial parsing) and it obtained an accuracy of 76.8% when evaluated on two subsets (1,677 pronouns) from a corpus of a telecommunications handbook and the Lexesp corpus (made up of newspaper articles and narratives). Saiz-Noeda [18] improves the system by incorporating syntactic functions –which allows a revision and optimization of the constraints and preferences of the original algorithm– as well as semantic information –WordNet synsets measure the degree of semantic compatibility between the antecedent and the verb. In its best performance on a 3,000-word fragment (two opinion articles and a narrative text) from the Lexesp corpus, Saiz-Noeda [18] reports an accuracy of 94.49%. His evaluation was on a small corpus from a closed domain. It is not clear if the system generalizes to open-domain corpora and to the many types of coreferential relations we propose in this paper.

The knowledge-based approach is the one that has dominated anaphora resolution for a long time. However, since the 90s, in order to cater for the processing of unrestricted corpora –essential in the Internet field–, there has been a growing need for wide coverage systems. In this context, the machine-learning-based approach may be better suited than rule-based coreference anaphora resolvers. Some systems have tried using non-annotated corpora, but some linguistic issues –such as anaphora resolution– require annotated data, as little can be learnt from raw texts. We aim at testing the success of a learning-based coreference system trained on an annotated 500,000-word corpus.

With respect to corpus-based techniques, as far as we know, for Spanish there is no substantial corpus available in which coreferential or anaphoric relations are encoded. Besides, all the research on anaphora resolution carried out up to now has focused either on pronouns or on DDs, but no project has dealt with pronouns, full NPs and proper nouns all together as we do. In order to enrich the Spanish CESS-ECE corpus with coreference information, we draw on projects developed for English, considering the markup schemes, tools and strategies that have been

suggested, and making the adaptations, changes and extensions that we feel necessary given the conditions of the corpus and our purposes.

## 4 Coreference annotation scheme

The CESS-ECE corpus is a multilingual corpus that consists of a Spanish (CESS-ESP) and a Catalan corpus (CESS-CAT), 500,000 words each mostly coming from newspaper articles. The CESS-ECE corpus has already been annotated with morphological information (PoS), syntactic constituents and functions, argument structures and thematic roles, tagged with strong and weak named entities (NE), and the 150 most frequent nouns have their WordNet synset [10]. It is the largest annotated corpus of Spanish. The information already annotated is taken into account when planning the enrichment of the corpus with coreference links.

The annotation methodology of the CESS-ECE corpus is divided into two steps: a first automatic stage, and a second manual one. The former takes advantage of the annotation already contained in the corpus; while the latter enriches manually the automatic annotation and incorporates the anaphoric and coreferential links.

With regard to the annotation scheme, after considering different ones [5, 7, 8, 12, 15, 19], we opted to implement the scheme proposed in MATE/GNOME [16] for coreference annotation, because its great flexibility and modularity make it able to meet our needs. It is open to linguistic phenomena of languages other than English and, although designed for dialogues, it can be easily adapted to other textual genres. Besides, it keeps distinct the annotation of discourse entities from the annotation of links.

A number of coreference annotation projects have drawn on the MUC-6 and MUC-7 schemes [7], in which two NPs are considered to be coreferential if they refer to the same entity in the world. However, van Deemter & Kibble [20] have criticized the MUC Task Definition for violating the relation of coreference proper and mixing it with anaphora. The MATE scheme differs from that of MUC in that it is based on the discourse rather than the world. Following the discourse model [21], coreference and anaphora occur between discourse entities (DE), which may or may not refer to specific objects in the world. So the first stage in the development of an annotation scheme is the delimitation of which text constituents realize DEs that may enter in coreferential relations and are identified as markables with a <de> element.

### 4.1 First step: Automatic annotation

The starting point is the rich hierarchical syntactic annotation contained in the CESS-ECE corpus (see Figure 2). The general tag `sn` codes all NPs, while more specific tags are used to mark coordinated NPs (`sn.co`), adjunct NPs (`sn.j`), NPs containing a coordinated nominal group (`sn.x`), and elliptical subjects (`sn.e`). All these tags are also able to contemplate cases of discontinuities (splitting into a `sn.1n` and a `sn.1c` labels) and they contain a further specification –an additional letter at the end of the tag– if they are NEs: `o` for organizations, `l` for locations, `p` for

persons, `d` for dates, `n` for numbers (including percentages and money), and `a` for the rest.

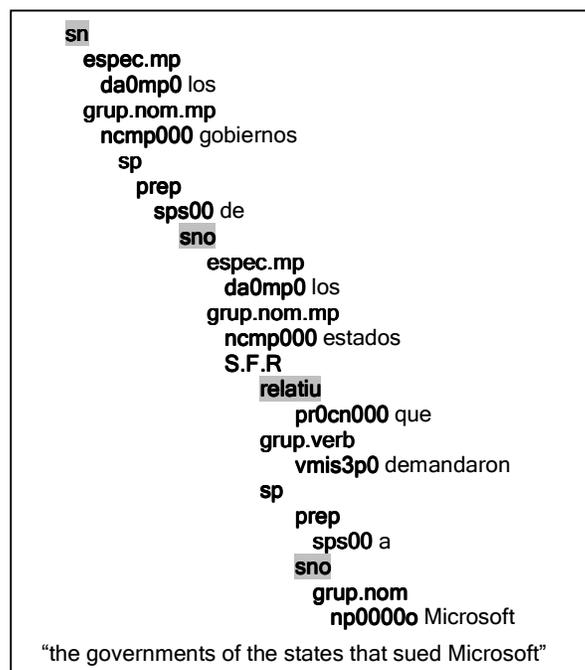


Figure 2: Fragment of a morphosyntactic tree from the CESS-ECE corpus

#### 4.1.1 Markables: <de> elements

As specified in Section 2, we aim at coding coreferential relations involving at least one NP –whose antecedent may be another NP, a VP, a clause or a sequence of clauses. Therefore, these are the text constituents that should be marked up. All NPs are automatically marked up, whether definite, indefinite, pronominal, bare nouns or proper names. Antecedents expressed by phrases other than nominal are later marked manually when necessary. The full syntactic annotation of the CESS-ECE corpus enables us to follow the MATE’s guidelines and do most of the markable identification task automatically, by instructing the computer to mark all `sn` (NPs) as <de> with an ID number, as shown by the highlighted nodes in Figure 2. Although not all NPs should be treated as markables, in this first automatic step no distinction is made. `Sn` syntactic tags treat relative clauses and appositional phrases as modifiers of the head noun, so both are included within the <de> tag. Although the Spanish reflexive pronoun *se* can also function as a verbal morpheme or as a mark for passive and impersonal constructions, the syntactic annotation already contains this information, so that the automatic annotation only identifies as <de> the uses that are really reflexive and so coreferential.

The fact that coordinated NPs, apart from their own tag, are syntactically marked with a tag for the larger NP means that three or more <de> are generated: one for each constituent NP and one for the larger NP. Subsequent references either to parts or to the whole coordination imply then no additional difficulty. Since relative

pronouns are tagged as *relatiu*, the computer is also instructed to mark them as `<de>`.

Unlike English, subject pronouns are usually omitted in Spanish –as they can be easily recovered from the verbal morphology. Otherwise a contrastive value is implied. So it is a great advantage that zero subject pronouns are syntactically already shown as \*0\*. It means that they can be automatically marked up as `<de>` and no other special tag is required.

On the other hand, in order to specify the antecedent of what we call “contextual descriptions”, we adapt the MATE’s possibility of annotating references to visible objects. Each piece of news is introduced by a `<universe>` element containing two universe entities (`<ue>`): the location and the time in which the piece of news was written (3). Both elements are automatically filled with the information heading each file.

(3) `<universe> <ue type="location" id="ue_1"> Toledo </ue> <ue type="date" id="ue_2"> 23.07.97 </ue> </universe>`

#### 4.1.2 Markables: the TYPE attributes

Apart from its ID, each `<de>` element has one or two TYPE attributes: the first specifies the type of NP –its degree of determination–, whereas the second appears only if the `<de>` is a NE or a self-sufficient DD. TYPE1 can be filled automatically by profiting from the morphological annotation of the corpus, thus copying the information contained in the specifier section (*espec.*), if there is any. In Figure 2, for instance, TYPE1 for the first `<de>` is filled with “da0mp0” (namely, determiner, article, masculine, plural). If the NP has no specifier, the information for TYPE1 is provided in the nominal group (*grup.nom*) section or in the node tag itself: “rel” for relative pronouns, “co” for NPs containing coordination, “e” for elliptical subjects (in this case, the verbal morphology is included as well), etc.

The TYPE2 attribute is automatically filled for NEs, whereas the identification of self-sufficient DDs is done in the manual annotation stage. After running the automatic annotation, the `<de>` elements obtained from the tree in Figure 2 are shown in (4).

(4) `<de id="de_0" type1="da0mp0"> los gobiernos de <de id="de_1" type1="da0mp0" type2="NE-org"> los estados <de id="de_2" type1="rel"> que </de> demandaron a <de id="de_3" type1="np00000" type2="NE-org"> Microsoft </de> </de> </de>`

## 4.2 Second step: Manual annotation

At this point two tasks need to be carried out. On the one hand, the automatic identification of markables is completed by adding unidentified ones. On the other hand, annotators have to annotate manually coreferential relations by incorporating the `<link>` element wherever necessary.

### 4.2.1 Adding markables

First, since incorporated clitics are not syntactically annotated in the corpus, annotators have to mark the verbal complex as a `<clit>` element (5), including as many `<clit>` as clitics there are.

(5) `<clit id="clit_12" type1="pp3cn000"> <clit id="clit_13" type1="pp3cna00"> dárselo </clit> </clit>`<sup>4</sup>

Second, antecedents corresponding to a VP, a clause or a sequence of clauses are marked as `<seg>` elements.

Third, the TYPE2 attribute needs to be filled with the value “SD” for DDs which are considered to be self-sufficient, that is, NPs with the definite article that depend on no antecedent, but on world knowledge. Their autonomy can result from their generic reference (6a), their containing an explanatory modifier (6b, 6c) or their general uniqueness (6d).

(6) a. los alemanes  
b. las reservas de oro y divisas del Banco Central  
c. los estados que demandaron a Microsoft  
d. la policía<sup>5</sup>

Marking these DDs as “SD” can prove successful for a resolution system when learning to recognize definite NPs that, like proper nouns, can potentially be the first elements of a coreference chain.

On the other hand, annotators are advised to omit `<de>` elements which participate very rarely in coreferential relations, such as pronouns referring to an adjective, bound anaphors (within the scope of a quantifier), bare NPs with an attributive value, idiomatic expressions, and pronouns within fixed connectors.

### 4.2.2 Annotating coreference: <link>

The `<link>` elements serve to show coreferential relations holding between two discourse entities. This marking is especially useful for question answering, information extraction as well as text summarization. The ANCHOR attribute points to the ID of the antecedent. For the sake of simplicity, we do not distinguish between anaphora and cataphora, so that it is possible that the ANCHOR entity appears not before but after its related `<de>`. We agree to mark the closest antecedent, whether pronominal or not, as the ANCHOR.

The TYPE attribute of the `<link>` specifies the kind of coreferential relation and can take seven different values (the last three ones unique to our scheme):

(i) **type=“ident”** (identity)

The two `<de>` share the same discourse referent. It may involve a full NP and a pronoun (7a), a proper noun and a pronoun, a proper noun and a full NP (7b), two proper nouns, or two full NPs, which may share the same head (7c) or stand in a synonymy, hypernymy or hyponymy relationship (7d). We also treat as identity relations the resolution of first and second person pronouns in quoted speech, as once within a written discourse, deictic pronouns are interpreted in an anaphoric way (7e).

(7) a. El presidente boliviano y el jefe del partido de la oposición...ambos.  
b. Microsoft...la firma.

<sup>4</sup> (5) ‘give-him/her/them-it’

<sup>5</sup> (6) a. the Germans

b. the gold and currency reserves of the Central Bank

c. the states that sued Microsoft

d. the police

- c. la falta de mano de obra en Cataluña...esta falta de mano de obra.
- d. un grupo de adolescentes...el equipo.
- e. “yo sigo” – dijo el director general de Seat.<sup>6</sup>

(ii) **type=“dx”** (discourse deixis)

The antecedent of the NP is a VP, a clause (8a), or a sequence of clauses (8b, 8c) –be it an event, fact, or proposition. The difficulty of deciding the exact textual part that serves as antecedent –which can be considerably long– has been pointed out by van Deemter & Kibble [20] and Poesio [15]. Given the relevance of events in NLP tasks, it is important for discourse deixis to have specific guidelines about how the ANCHOR should be marked. These guidelines will appear in [17]. The resolution of discourse deixis helps answer fusion in question answering and template merging in information extraction.

- (8) a. Si no cambia la situación meteorológica, cosa que el INM no prevé a corto plazo,...
- b. Pujol cree necesario que el Gobierno agilice los permisos de residencia a los inmigrantes para... Esta opinión...
- c. [...] Con esto no quiero decir que nosotros...<sup>7</sup>

(iii) **type=“poss”** (possessor)

The possessor link concerns possessive pronouns, NPs introduced by a possessive determiner, and possessive relatives. The coreference relation shows that the <de> antecedent is the possessor of the second <de>, which may express an object properly possessed as well as a part or an attribute of the possessor (9). Unlike cases of part-of bridging, possessor relations are straightforward, as the possessive makes them explicit.

- (9) El primer ministro mostró su preocupación.<sup>8</sup>

(iv) **type=“bridg”** (bridging)

This is a very broad class that encompasses all kinds of metonymic relations –to a greater or lesser extent– holding between two NPs (subset, member, etc.) (10), or between a NP and a VP, implicitly related. Bridging is treated within coreference in the sense that the link between the two discourse entities is established on the basis of the same reference point. A detailed specification of bridging subtypes is addressed in [17].

- (10) a. el cambio de 17 acciones de Alcan...los accionistas
- b. la tropa...uno de los soldados.<sup>9</sup>

<sup>6</sup> (7) a. The Bolivian president and the head of the opposition party...both.  
b. Microsoft...the firm.  
c. the lack of labour in Catalonia...this lack of labour.  
d. a group of adolescents...the team.  
e. “I go on” – said the general director of Seat.

<sup>7</sup> (8) a. If the meteorological situation does not change, something that the INM does not foresee in the short term...  
b. Pujol believes it necessary that the government speeds up the residence permits for immigrants to...This opinion...  
c. [...] With this I do not want to say that we...

<sup>8</sup> (9) The Prime Minister showed his concern.

(v) **type=“pred”** (predicative)

Following van Deemter & Kibble [20], we do not treat nominal predicates (11a) and appositional phrases (11b) as identity coreference. However, given that NPs identifying a discourse entity by its properties can be very relevant for some NLP tasks –such as Entity Detection and Recognition from ACE, and definitional question answering–, we have created the special “predicative link” type for these cases.

- (11) a. Villatoro es el director del diario Avui.
- b. Barnasants, el ciclo de canción de autor...<sup>10</sup>

(vi) **type= “rank”** (ranking)

The ranking link applies to NPs that refer to the numerical order of the elements of a given list. The ANCHOR is either a coordinated or a complex NP of the enumerative kind (12). This link helps “list” questions in question answering, e.g. “Name all the participants in the event.”

- (12) Por este orden, participaron en el acto Javier Krahe, Javier Ruibal y Loquillo.<sup>11</sup>

(vii) **type=“context”** (contextual)

Contextual descriptions are interpreted with respect to the spatial or temporal coordinates (13). Therefore, their ANCHOR is not a <de>, but one of the two universe entities from the <universe> element.

- (13) Este año las cifras están por debajo de la media.<sup>12</sup>

When considering the taxonomy of coreferential link types, we decided to include a second kind of <link> element –different from the coreferential one– so as to fill the semantic type of headless NPs. The <sem type:link> element (with an ANCHOR attribute) is limited to some NPs with adjectives (14a), PPs (14b) or relative clauses as heads.

- (14) a. Tres tipos de vestidos: los blancos, los...
- b. Hubo poca participación, pero la de los españoles...<sup>13</sup>

## 5 Conclusions and further work

In this paper we have presented a foundational step for the annotation of the CESS-ECE corpus with coreference information: the design of the guidelines and general criteria to carry out our project. This annotation scheme allows us to annotate a corpus sample and identify problems and unexpected cases that lead us to extend and refine the markup scheme. Therefore, the scheme here presented is open to new attributes and values. The outcome of this long process is the definition of an

<sup>9</sup> (10) a. the change of 17 shares of Alcan...the shareholders  
b. the troop...one of the soldiers.

<sup>10</sup> (11) a. Villatoro is the director of the Avui newspaper.  
b. Barnasants, the singer-writer song cycle...

<sup>11</sup> (12) In this order, took part in the event Javier Krahe, Javier Ruibal and Loquillo.

<sup>12</sup> (13) This year the figures are below the mean.

<sup>13</sup> (14) a. Three types of dresses: the white ones, the...  
b. There was little participation, but that of the Spanish...

exhaustive typology of coreferential expressions in Spanish<sup>14</sup> and coreference relations.

In contrast to existing anaphora resolution systems for Spanish, our project covers the whole range of coreferential expressions, thus dealing with proper nouns, full NPs and pronouns all together. The existence of a rich syntactic annotation in the CESS-ECE corpus offers the possibility of doing most of the markable identification automatically, thus reducing the extent of manual annotation, which is well known to be a labour-intensive and time-consuming task.

The choice of the annotation tool is another key point, and we are considering the way in which the existing ones can be adapted to meet our needs. Regarding the annotation strategy, annotators meet periodically to discuss the doubtful cases and thus achieve a level of inter-annotator agreement as high as possible.

This work is the first step in the creation of the largest corpus with complex semantic annotation for Spanish. Once the CESS-ECE corpus is annotated following a scheme linguistically well founded, the goal of our project is twofold. From a computational perspective, the development of an automatic coreference resolution system by applying machine-learning techniques. Besides, the annotated corpus can be used by researchers to train and test automatic coreference resolution methods. From a linguistic point of view, we shall test the hypotheses suggested by Ariel [1] and Gundel et al. [6] on the basis of the annotated data in the CESS-ECE corpus. The linguistic study may lead us to infer generalisations about the expression of coreference in Spanish that can be used as heuristics. According to Botley & McEnery [4], the existing variety of approaches and methodologies to anaphora resolution calls for a synthesis. The combination of the machine-learning algorithms with the heuristics obtained from the linguistic analysis will be a fruitful synthesis, resulting in a hybrid coreference resolution system.

## Acknowledgements

We would like to thank Mihai Surdeanu for his helpful advice and suggestions.

This paper has been supported by the FPU grant (AP2006-00994) from the Spanish Ministry of Education and Science. It is based on work supported by the CESS-ECE (HUM2004-21127), Lang2World (TIN2006-15265-C06-06), and Praxem (HUM2006-27378-E) projects.

## References

- [1] M. Ariel. Referring and accessibility. *Journal of Linguistics*, 24(1):65-87, 1988.
- [2] S. Blackwell. *Implicatures in Discourse: The Case of Spanish NP Anaphora*. John Benjamins, Amsterdam, 2003.
- [3] I. Bosque and V. Demonte (editors). *Gramática descriptiva de la lengua española*, Real Academia Española / Espasa Calpe, Madrid, 1999.
- [4] S. Botley and A. McEnery (editors). *Corpus-based and computational approaches to discourse anaphora*. John Benjamins, Amsterdam, 2000.
- [5] C. Gardent and H. Manuélian. Création d'un corpus annoté pour le traitement des descriptions définies. *Traitement Automatique des Langues (TAL)*, 46(1):115-140, 2005.
- [6] J. K. Gundel, N. Hedberg, and R. Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274-307, 1993.
- [7] L. Hirschman and N. Chinchor. MUC-7 coreference task definition. In *MUC-7 Proceedings*. Science Applications International Corporation, 1997.
- [8] V. Hoste and W. Daelemans. Learning Dutch coreference resolution. In *Proceedings of the 15<sup>th</sup> Computational Linguistics in the Netherlands Meeting (CLIN 2004)*, 2005.
- [9] S. Lappin and H. J. Leass. An algorithm for pronominal anaphora resolution, *Computational Linguistics*, 20(4):535-561, 1994.
- [10] M. A. Martí, M. Taulé, L. Márquez, and M. Bertran. CESS-ECE: a multilingual and multilevel annotated corpus, to appear. <http://www.lsi.upc.edu/~mbertran/cess-ece/publications>.
- [11] R. Mitkov. *Anaphora Resolution*. Longman, London, 2002.
- [12] R. Mitkov, R. Evans, C. Orasan, C. Barbu, L. Jones, and V. Sotirova. Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies. In *Proceedings of the 3<sup>rd</sup> Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lancaster, 49-58, 2000.
- [13] R. Muñoz. *Tratamiento y resolución de las descripciones definidas y su aplicación en sistemas de extracción de información*. Ph.D. Thesis, University of Alicante, Spain, 2001.
- [14] M. Palomar, A. Ferrández, L. Moreno, P. Martínez-Barco, J. Peral, M. Saiz-Noeda, and R. Muñoz. An algorithm for anaphora resolution in Spanish texts. *Computational Linguistics*, 27(4):545-567, 2001.
- [15] M. Poesio. MATE Dialogue Annotation Guidelines – Coreference. Deliverable D2.1, 2000. <http://www.ims.uni-stuttgart.de/projekte/mate/mdag>
- [16] M. Poesio. The MATE/GNOME proposals for anaphoric annotation, revisited. In *Proceedings of the 5<sup>th</sup> SIGdial Workshop on Discourse and Dialogue*, Boston, 154-162, 2004.
- [17] M. Recasens, M. A. Martí, and M. Taulé. Text as scene: discourse deixis and bridging relations. In *Proceedings of the Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN2007)*, Sevilla, 2007.
- [18] M. Saiz-Noeda. *Influencia y aplicación de papeles sintácticos e información semántica en la resolución de la anáfora pronominal en español*. Ph.D. Thesis, University of Alicante, Spain, 2002.
- [19] A. Tutin, F. Trouilleux, C. Clouzot, E. Gaussier, A. Zaenen, S. Rayot, and G. Antoniadis. Annotating a large corpus with anaphoric links. In *Proceedings of the 3<sup>rd</sup> Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lancaster, 2000.
- [20] K. van Deemter and R. Kibble. On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4):629-637, 2000.
- [21] B. Webber. *A Formal Approach to Discourse Anaphora*. Garland Press, New York, 1979.

<sup>14</sup> This work extends easily to other Iberian Romance languages such as Catalan and Galician.