Procesamiento del Lenguaje Natural, nº39 (2007), pp. 205-212 recibido 02-05-2007; aceptado 22-06-2007

Text as Scene: Discourse Deixis and Bridging Relations

Marta Recasens
Universitat de Barcelona
Gran Via Corts Catalanes,585
08007 Barcelona
mrecasens@ub.edu

M. Antònia Martí
Universitat de Barcelona
Gran Via Corts Catalanes,585
08007 Barcelona
amarti@ub.edu

Mariona Taulé
Universitat de Barcelona
Gran Via Corts Catalanes,585
08007 Barcelona
mtaule@ub.edu

Abstract: This paper presents a new framework, "text as scene", which lays the foundations for the annotation of two coreferential links: discourse deixis and bridging relations. The incorporation of what we call *textual* and *contextual scenes* provides more flexible annotation guidelines, broad type categories being clearly differentiated. Such a framework that is capable of dealing with discourse deixis and bridging relations from a common perspective aims at improving the poor reliability scores obtained by previous annotation schemes, which fail to capture the vague references inherent in both these links. The guidelines presented here complete the annotation scheme designed to enrich the Spanish CESS-ECE corpus with coreference information, thus building the CESS-Ancora corpus.
Keywords: corpus annotation, anaphora resolution, coreference resolution.

Resumen: En este artículo se presenta un nuevo marco, "el texto como escena", que establece las bases para la anotación de dos relaciones de correferencia: la deixis discursiva y las relaciones de *bridging*. La incorporación de lo que llamamos *escenas textuales* y *contextuales* proporciona unas directrices de anotación más flexibles, que diferencian claramente entre tipos de categorías generales. Un marco como éste, capaz de tratar la deixis discursiva y las relaciones de *bridging* desde una perspectiva común, tiene como objetivo mejorar el bajo grado de acuerdo entre anotadores obtenido por esquemas de anotación anteriores, que son incapaces de captar las referencias vagas inherentes a estos dos tipos de relaciones. Las directrices aquí presentadas completan el esquema de anotación diseñado para enriquecer el corpus español CESS-ECE con información correferencial y así construir el corpus CESS-Ancora.
Palabras clave: anotación de corpus, resolución de la anáfora, resolución de la correferencia.

1 Introduction

Due to the lack of large annotated corpora with anaphoric information, the field of computational coreference resolution is still highly knowledge-based, especially for languages other than English. 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 noun-phrase (NP) coreference information (thus building the CESS-Ancora corpus). The design of the annotation guidelines is presented in (Recasens, Martí & Taulé, 2007), but two types of coreferential links, namely discourse deixis[1] and bridging relations[2], call for a specific analysis which takes into account their complex peculiarities so as to determine the most appropriate set of attributes and values.

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 three series of demonstratives and pronouns marked for neuter gender. The typology presented in this paper is the completion of a flexible annotation scheme rich enough to cover the cases of coreference in Spanish.

[1] We define discourse deixis (or abstract anaphora) as reference to a discourse segment, that is, to a non-nominal antecedent.

[2] Our approach classifies as bridging (or associative anaphors) those definite or demonstrative NPs that are interpreted on the grounds of a metonymic relationship with a previous NP or VP.

ISSN: 1135-5948

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 (1988) and Gundel, Hedberg & Zacharski (1993) 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.

As a follow-up, this paper places the emphasis on the annotation guidelines for discourse deixis and bridging relations. Both are considered from a common perspective: what we call "text as scene", that is, the text taken as a scene in the sense that it builds up both a *textual* and a *contextual* framework as the result of an interaction between the discourse and the global context.

The rest of the paper proceeds as follows: Section 2 reviews previous work on abstract and bridging anaphora. A description of the "text as scene" framework is provided in Section 3. Specific guidelines for annotating discourse deixis and bridging relations are given in Section 4. Finally, Section 5 presents our conclusions and discussion of the guidelines.

2 Previous work

Given the difficulty of dealing with antecedents other than NPs, most of the work on anaphora resolution has ignored abstract anaphora and has limited to individual anaphora. However, the work of Byron (2002) has emphasized that demonstrative pronouns referring to preceding clauses abound in natural discourse[3]. In this line, the corpus-based study of the use of demonstrative NPs in Portuguese and French conducted by Vieira et al. (2002) has pointed out that a system limited to the resolution of anaphors with a nominal antecedent is likely to fail on about 30% of the cases.

In her seminal study, Webber (1988) coins the term "discourse deixis" for reference to discourse segments and argues that these should be included in the discourse model as discourse entities, since they can be subsequently referenced via deictic expressions. Nevertheless, a discourse entity corresponding to a textual segment is not added to the discourse model until the listener finds a subsequent deictic pronoun, in the so-called *accommodation* process[4]. Works on parsing texts into discourse segments (Marcu, 1997) have not dealt with the problem of discourse deixis, i.e. delimiting the extent of the antecedent.

With respect to corpus annotation, there are not many annotation schemes that annotate antecedents other than NPs. The MUC Task Definition (Hirschman & Chinchor, 1997) explicitly defines demonstratives as non-markables. Two notable exceptions are the MATE scheme by Poesio (2000) and the scheme by Tutin et al. (2000), although both point out the difficulty of delimiting the exact part of the text that counts as antecedent as well as the type of object the antecedent is. Tutin et al. (2000) decide to select the largest possible antecedent.

Similarly to discourse deixis, authors seem sceptical about the feasibility of the annotation task for bridging relations, especially since the empirical study conducted by Poesio & Vieira (1998), which reported an agreement of 31%. The issue under debate is where the boundary lies between a discourse-new NP and a bridging one, that is, between autonomous and non-autonomous definite NPs. Fraurud's (1990) starting point for her corpus-based study is a two-way distinction between first-mentions and subsequent mentions (coreferential NPs). On realising that 60% of the definite NPs were first-mention uses, she concludes that in addition to the syntactic (in)definiteness of an NP, the lexico-encyclopaedic knowledge associated with the head noun of the NP interacts with the general knowledge associated with present anchors in order to select one or more anchors in relation to which a first-mention definite NP is interpreted. Anchors may be provided in the discourse itself –either explicitly or implicitly–, by the global context, or by a combination of the two. Although Fraurud does not use the term, the first-mention NPs that are interpreted in relation to an explicit anchor correspond to "bridging relations".

[3] Byron's anaphora resolution algorithm differentiates Mentioned Entities (those evoked by NPs) from Activated Entities (those evoked by linguistic constituents other than NPs, involving global focus entities).

[4] Accommodation results from the use of a singular definite, which is felt to presuppose that there is already a unique entity in the context with the given description that will allow a truth value to be assigned to the utterance (Lewis, 1979).

In their analysis of the use of pronouns and demonstrative NPs in bridging relations, Gundel, Hedberg & Zacharski (2000) conclude that such cases are best analysed as minor violations to the Giveness Hierarchy, in that the listener gets away with an underspecified referent on the basis of what is predicated in the text.

What do then discourse deixis and bridging relations have in common? On the one hand, they are the anaphoric links with poorest reliability scores. On the other hand –and probably a cause of the former–, their antecedents are rather fuzzy, either because their extension cannot be clearly determined or because the semantic relation that links them with their anaphor cannot be easily identified. Taking into account the low inter-annotator agreement together with the idea of vague reference, we propose viewing the text as a scene in order to provide a wider contextual framework that captures those cases in which a discourse entity alludes to something that is not literally mentioned in the discourse.

3 Text as scene

Previous aims at annotating coreference have shown the need for reconsidering the annotation of both discourse deixis and bridging relations, since the reference of NPs such as *esto*, *la cosa*, and *este mercado* in (1), (2) and (3) respectively[5] cannot be accounted for from approaches that insist on linking each anaphoric expression to an explicit textual antecedent.

(1) El Komercni Banka –Banco Comercial–, uno de los cuatro bancos más grandes de la República Checa, anunció hoy que despedirá a 2.300 empleados más antes de finales del año dentro del proceso de saneamiento de la entidad estatal. El director del banco, Radovan Vrava, señaló que el motivo principal es la reestructuración del banco. El Estado dispone del 60 por ciento de las acciones del Komercni Banka y el Gobierno checo quiere comenzar el proceso de privatización de este banco ya en este año y terminarlo en septiembre del 2001. Otro de los objetivos es evitar que se repitan los errores del pasado, que obligaron al Gobierno a comprar créditos dudosos por un valor de 60.000 millones de coronas –1.500 millones de dólares. Esto permitirá al banco sanear su portafolio...[6]

(2) "Las previsiones para los próximos diez días no son nada halagueñas", pronosticó ayer Eduardo Coca, director del Instituto Nacional de Meteorología. Tan sólo un pequeño frente con poca agua debía cruzar el norte de la península entre ayer y hoy. Por lo demás, seguirá la situación anticiclónica. Pero la cosa no acaba ahí.[7]

(3) El presidente de la Comisión del Mercado de las Telecomunicaciones mostró su preocupación por la falta de competencia en *la telefonía local*, como consecuencia de que la liberalización de las telecomunicaciones se ha hecho por principios jurídicos y no técnicos y que "hay que abrir este mercado como sea".[8]

[5] The reader is asked to please forgive the length of most of the examples used in this paper, but the anaphoric expressions we deal with make no sense unless the context is provided.

[6] (1) The Komercni Banka –Commercial Bank –, one of the four biggest banks in the Cheque Republic, announced today that it will dismiss 2,300 more workers by the end of the year within the reform process of the state entity. The director of the bank, Radovan Vrava, pointed out that the main reason is the restructuration of the bank. The State possesses the 60 per cent of the shares of the Komercni Banka and the Cheque Government wants to begin the privatisation process of this bank already this year and finish it in September 2001. Another of the goals is to avoid the repetition of past mistakes, which forced the Government to buy doubtful credits for the price of 60,000 million crowns –1,500 million dollars. This will allow the bank to reform its portfolio.

[7] (2) "The forecasts for the next ten days are not favourable at all", forecasted yesterday Eduardo Coca, director of the National Institute of Meteorology. Only a small front with little water should cross the north of the peninsula between yesterday and today. As for the rest, the anticyclonic situation will persist. But the thing does not end there.

[8] (3) The president of the Commission of the Market of Telecommunications showed his concern for the lack of competence in *local telephony*, as a

Our coding scheme is defined from the consideration of the text as a scene in two different senses (see Figure 1), the scene being the cohesive element. On the one hand, discourse deixis captures those anaphoric expressions that refer back to the *textual scene*, that is, to a discourse segment –either at the sentence level or beyond the sentence– that builds up a scene as a whole. On the other hand, bridging captures those implicit relations (between two discourse entities) that are enabled by the *contextual scene* activated by the involved entities. A contextual scene is taken to be the knowledge which does not explicitly appear in the text, but that contributes to its comprehension. Bridging is treated within coreference in the sense that the two discourse entities share the reference point on the basis of a contextual scene.

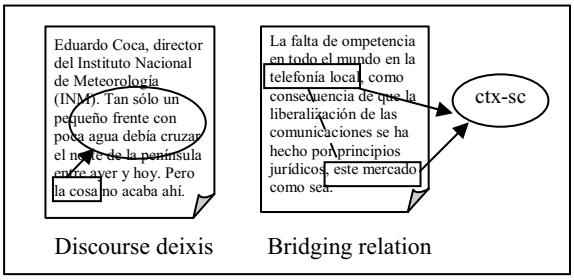

Figure 1: Textual and contextual scenes

Back to example (1), the discourse segment picked up by the pronoun *esto* –that which is going to allow the Cheque Bank to reform its portfolio– results not only from the last discourse segment, but from combining the content of the events that form the entire textual scene: the dismissal of 2,300 workers, the restructuration of the Bank, its privatisation, and the avoidance of past mistakes. Similarly, the definite NP *la cosa* in (2) makes reference to the textual scene previously described. It becomes a quasi-pronominal form in that it is almost semantically empty. Finally, example (3) shows a case of bridging, where the interpretation of the demonstrative NP *este mercado* is made possible by the contextual scene activated by a former NP, *la telefonía local*, namely, the market opened by local telephony.

Text as scene provides a common framework within which we are able to reach a consensus as to the typology of referring expressions that can code discourse deixis and bridging relations as well as the subtypes of links that need to be annotated with a view to achieving a level of inter-annotator agreement as high as possible.

4 Corpus annotation

The CESS-ECE corpus is the largest annotated corpus of Spanish, which contains 500,000 words mostly coming from newspaper articles. It has been annotated with morphological information (PoS), syntactic constituents and functions, argument structures and thematic roles, tagged with strong and weak named entities, and the 150 most frequent nouns have their WordNet synset.

Drawing from the MATE scheme (Poesio, 2000) and taking into account the information already annotated, the enrichment of the corpus with coreference annotation is divided into two steps: a first automatic stage, and a second manual one. The former marks up all NPs of the corpus as <de> (discourse entity) with an ID number, and fills in the TYPE attributes with morphological information (the kind of NP); the latter step adds those <de> unidentified by the automatic annotation – and codes the coreferential relations by incorporating the <link> element.

It is at this second stage when antecedents expressed by phrases other than nominal are marked manually as <seg> elements when necessary. The <coref:link> elements serve to show coreferential relations holding between two discourse entities, and the embedded <coref:anchor> element points to the ID of the antecedent. Each <coref:link> has a TYPE attribute that specifies the kind of coreferential relation. We distinguish seven types of links:

(i) ident (identity)
(ii) dx (discourse deixis)
(iii) poss (possessor)
(iv) bridg (bridging)
(v) pred (predicative)
(vi) rank (ranking)
(vii) context (contextual)

Given that the marking of both discourse deixis and bridging relations is very useful for tasks such as question answering (answer fusion), information extraction (template merging) and text summarization, but that the annotation of these two links poses great difficulty, we

consequence of the fact that the liberalisation of telecommunications has been done by juridical and not technical principles and that "this market must be opened at all costs".

consider it necessary to devote the two following sections to specifying their annotation guidelines, which are based on our conception of the text as scene.

4.1 Discourse deixis (dx)

We consider an anaphoric NP to be in a dx relation when its antecedent is a textual scene expressed by a clause or a sequence of clauses. NPs that have the potential to participate in dx links are demonstrative pronouns, the neuter personal pronoun *lo*, the relative pronoun *que*, demonstrative full NPs, and definite descriptions (DD) of the kind *la cosa, el fenómeno, la situación*, etc. We call these NPs "quasi-pronominal DDs", as they can be replaced by the pronoun *esto* and are almost empty of semantic content of their own.

Textual scenes are not constituted as such until a corresponding referring expression appears in the discourse. The pronouns *lo* and *que* tend to refer to textual scenes within the same discourse segment or introduced in the previous sentence, while demonstratives and quasi-pronominal DDs can refer to scenes that are more than one sentence away. Since it is not a trivial matter to decide the exact part of the text that serves as antecedent, we distinguish between two SUBTYPE attributes for dx:

(i) subtype="sent" (sentential)

This subclass covers the less problematic cases of discourse deixis, i.e. those anaphoric NPs that refer to a textual scene whose extent is no longer than one sentence (any discourse segment from period to period). We mark the non-nominal antecedent as a <seg> element with an ID number, which fills the <coref:anchor>. When in doubt about the exact delimitation of the text segment, the entire sentence is marked-up. For ease of presentation, (4a) shows the extent of the antecedent for the anaphoric demonstrative NP *este camino*[9], whereas (4b) codes the link as it is done in the annotation of the CESS-Ancora corpus.

Taking into account that the pronoun alone is not enough to pick up its referent, but that this is made clear from the predicate complement information (Byron, 2000), we further determine the "sent" value with the semantic type of the antecedent: "sent-ev" for events (4), "sent-fact" for facts (5), and "sent-prop" for propositions (6).

(4) a. La ministra Anna Birulés animó a las pymes a **[**invertir en Investigación y Desarrollo**]** y *0* mostró a los empresarios presentes la disposición del Gobierno a facilitar este camino.[10]

b. La ministra Anna Birulés animó a las pymes a **<seg ID="seg_03">** invertir en Investigación y Desarrollo **</seg>** y *0* mostró a los empresarios presentes la disposición del Gobierno a facilitar **<de type="dd0ms0" ID="de_06">** este camino **</de>**. **<coref:link ID="de_06" type="dx" subtype="sent-ev"> <coref:anchor ID="seg_03"/> </coref:link>**

(5) Sin embargo, **[**los virus logran poner a su servicio al organismo vivo más desarrollado que existe: el ser humano.**]** Es éste un hecho que hace temblar el edificio que la humanidad ha construido.[11]

(6) **[**La Coordinadora de Organizaciones de Agricultores y Ganaderos teme que la falta de lluvia afecte también a los regadíos, dado que empieza a reducirse el volumen de agua embalsada.**]** Este temor es compartido por...[12]

(ii) subtype="text" (textual scene)

The textual scene subtype includes those cases discussed in Section 3 ((1) and (2)), where an anaphoric expression refers to the whole scene built up by the preceding text. These are cases that result from global discourse effects, so the precise anchor goes beyond the single sentence level and is usually vague in reference.

[9] In the examples, underlines correspond to anaphoric expressions, while square brackets identify their antecedents.

[10] (4) The minister Anna Birulés stimulated the SMEs **[**to invest in Research and Development**]** and showed the present businessmen the Government's willingness to facilitate this path.

[11] (5) Nevertheless, **[**viruses manage to put at their service the most developed living organism that exists: the human being.**]** This is a fact that makes the edifice that humanity has built tremble.

[12] (6) **[**The Coordinator of Organisation of Farmers fears that the lack of rain also affects irrigations, given that the volume of dammed water is starting to decrease.**]** This fear is shared by...

Therefore, as <coref:anchor> we indicate the ID of the paragraph (<par>) to which the anaphor belongs, thus indicating that the reference is made to the textual scene going from the beginning of the paragraph to the anaphor. As example, (7) shows the annotation for the anaphoric NP in (1).

(7) **<de type="pd0ns00" ID="de_09">** Esto **</de>** permitirá al banco sanear su portafolio.[13]
<coref:link ID="de_09" type="dx" subtype="text" > <coref:anchor ID="par_05"/> </coref:link>

Demonstratives which are part of idiomatic phrases, such as the connectors *de esta forma* or *en este sentido*, are not considered as markables, since they are mere linking phrases.

4.2 Bridging relations (bridg)

Bridging relations only make sense if we understand them as occurring within the contextual scene triggered by the interaction between two discourse entities. The set of bridging relations is still an open issue (see the classification schemes of Clark, 1977; Vieira, 1998; Poesio, 2000; Muñoz, 2001; Gardent, Manuélian & Kow, 2003), since rather than a binary distinction between first-mention and bridging NPs, there is a scale ranging from those definite NPs which are uniquely interpretable by means of world knowledge (i.e. self-sufficient definite descriptions (SD)[14]) to those definite NPs which depend on a previous anchor. Inevitably, however, many real examples remain in between, as in (8), where *todas las administraciones* does not mean "all administrations" (in the world), but just the subset relevant to this scene.

(8) La última edición de Barnasants, el ciclo de canción de autor, ha atraído, según su director, Pere Camps, a unas 2.000 personas. Camps destaca el apoyo unánime de todas las administraciones en la edición de este año.[15]

In our annotation scheme, we consider NPs such as that in (8) as generic. They are framed by the textual scene, but do not require any anchor for their interpretation. Therefore, first-mentions of such NPs are considered to be SDs, while subsequent mentions are annotated as identity coreference.

We limit the term bridging to NPs (either definite or demonstrative) that are metonymically interpreted –to a greater or lesser extent– on the basis of a former NP or VP. The second discourse entity is anchored on the entity which contributes to activating the necessary scene for its interpretation. Within the "text as scene" approach, all bridging relations are taken to be *contextual scene relations*. So we only subspecify three very basic distinctions, which tend to be widely agreed upon. The three SUBTYPE attributes are:

(i) subtype="part" (part-of)
The antecedent of the anaphoric NP corresponds to the whole of which the anaphor is a part, as in (9).

(9) La reestructuración de **[**los otros bancos checos**]** se está acompañando por la reducción del personal.[16]

(ii) subtype="member" (set-member)
As illustrated by (10), the subsequent NP refers to one or more members of the set expressed by the NP anchor.

(10) a. **[**la tropa**]**...uno de los soldados.

b. Ante **[**unas mil personas**]**, entre ellas la ministra de Ciencia y Tecnología, Anna Birulés, el alcalde de Barcelona, Joan Clos, la Delegada del Gobierno, Julia García Valdecasas, y una representación del gobierno catalán, Pujol dijo...[17]

[13] (7) This will allow the bank to reform its portfolio.

[14] We consider as SD those NPs with the definite article that depend on no antecedent, but on world knowledge. Their autonomy can result from their generic reference, their containing an explanatory modifier, or their general uniqueness.

[15] (8) The last edition of Barnasants, the singer-writer song cycle, has attracted, according to its director, Pere Camps, about 2,000 people. Camps emphasizes the unanimous support of all the administrations in the edition of this year.

[16] (9) The restructuration of **[**the other Cheque banks**]** is accompanied by the reduction of the staff.

[17] (10) a. **[**the troop**]**...one of the soldiers.

b. Before about **[**one thousand people**]**, among them the minister of Science and Technology, Anna Birulés, the mayor of Barcelona, Joan Clos, the Delegate of the Government, Julia García Valdecasas, and a representation of the Catalan government, Pujol said...

(iii) subtype ="them" (thematic)
The anaphoric NP is related to a VP (the anchor) via a thematic relation. In (11), for instance, *estas inversiones* is the patient of the previous verb *invertir*. Like sentential anchors in discourse deixis, antecedents corresponding to VPs are marked by hand with a <seg> tag.

(11) *0* podría hacer que la empresa dominante dejara de **[**invertir en la red**]** por no considerarla como una inversión atractiva, y el Gobierno debe incentivar estas inversiones.[18]

If no subtype is specified, it means that the anaphoric NP is interpreted on the basis of a contextual scene, but that it is not related to its anchor via a clear part-of, set-member or thematic relation. This includes cases commonly referred to as "discourse topic" or general "inference" bridging. Examples can be found in (3) and (12).

(12) El cambio de **[**17 acciones de Alcan**]**...los accionistas.[19]

5 Conclusions and discussion

In this paper we have developed the specific framework, "text as scene", on which we base the annotation guidelines for both discourse deixis and bridging relations. The former is annotated as coreferring with a certain *textual scene*, while the latter is coded on the basis of a *contextual scene* activated by the conjunction of two discourse entities.

Given the rather vague antecedents that anaphoric expressions interpreted via either of these relations have, the annotation of both discourse deixis and bridging relations has usually obtained considerably low inter-annotator agreement. Our annotation scheme is unique in that we deal with these two relations from a common framework. In contrast to other annotation schemes, ours assumes two additional sources for the referent to be interpreted –a textual and a contextual scene–, which allow broader categories and thus more flexible annotation guidelines. Other interesting contributions of our scheme are the consideration of what we call "quasi-pronominal DDs" as discourse deictics together with the inclusion of demonstrative NPs into the range of potential candidates for bridging relations.

These guidelines complete the annotation scheme designed to enrich the Spanish CESS-ECE corpus with coreference information, thus giving birth to the CESS-Ancora corpus. It is a scheme rich enough to cover the different types of coreference in Spanish. Nevertheless, coreference annotation is such a complex task – involving several types of linguistic items and different factors responsible for linking two items as coreferential– that we are currently conducting a reliability study on a subset of the corpus to investigate the feasibility and validity of our annotation scheme. The results obtained might lead us to extend and refine it. One of the issues whose reliability needs to be proved is the extent to which abstract antecedents can be semantically classified into events, facts and propositions.

We believe that a 500,000-word corpus annotated from the morphological to the pragmatic level may shed new light on key factors about the nature and working of expressions creating coreference. It has not been determined yet, for instance, the way contextual scenes come into play or their scope (Fraurud, 1990). The CESS-Ancora corpus will provide quantitative data from natural written discourse from which it will be possible to infer more precise and realistic linguistic generalisations about the use of coreferential and anaphoric expressions in Spanish.

On the other hand, the rich tagset that distinguishes seven types of coreferential relations and that separates individual from abstract anaphora (each with different attributes) makes the CESS-Ancora corpus a very fruitful language resource. Being publicly released, it shall be used both for training and evaluating coreference resolution systems, as well as in competitions such as ACE or ARE.

In brief, the goal of our project is twofold. From a computational perspective, the CESS-Ancora corpus will be used to construct an automatic corpus-based coreference resolution system for Spanish. From a linguistic point of view, hypotheses on the use of coreferential expressions (Ariel, 1988; Gundel et al., 1993) will be tested on the basis of the annotated data and new linguistic theories might emerge.

[18] (11) S/he could make the dominant company stop **[**investing in the net**]** for not considering it as an attractive inversion, and the Government must motivate these inversions.

[19] (12) The change of **[**17 shares**]** of Alcan...the shareholders.

Acknowledgments

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

Ariel, M. 1988. Referring and accessibility. *Journal of Linguistics*, 24(1):65-87.

Byron, D. K. 2000. Semantically enhanced pronouns. In *Proceedings of the 3rd Discourse Anaphora and Anaphor Resolution Colloquium (DAARC2000)*, Lancaster.

Byron, D. K. 2002. Resolving pronominal reference to abstract entities. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, Philadelphia, 80-87.

Clark, H. 1977. Bridging. In P.N. Johnson-Laird and P.C.Wason (editors), *Thinking: Readings in Cognitive Science*, Cambridge University Press.

Fraurud, K. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7:395-433.

Gardent, C., H. Manuélian, and E. Kow. 2003. Which bridges for bridging definite descriptions? In *Proceedings of the EACL 2003 Workshop on Linguistically Interpreted Corpora*, Budapest, 69-76.

Gundel, J., N. Hedberg, and R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274-307.

Gundel, J., N. Hedberg, and R. Zacharski. 2000. Statut cognitif et forme des anaphoriques indirects. *Verbum*, 22:79-102.

Hirschman, L. and N. Chinchor. 1997. MUC-7 coreference task definition. In *MUC-7 Proceedings*. Science Applications International Corporation.

Lewis, D. 1979. Score keeping in a language game. In R. Bäuerle et al. (editors), *Semantics from a different point of view*. Springer Verlag, Berlin.

Marcu, D. 1997. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. PhD Thesis, Department of Computer Science, University of Toronto.

Muñoz, R. 2001. *Tratamiento y resolución de las descripciones definidas y su aplicación en sistemas de extracción de información*. PhD Thesis, Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante.

Poesio, M. 2000. MATE Dialogue Annotation Guidelines – Coreference. Deliverable D2.1. http://www.ims.uni-stuttgart.de/projekte/mate/mdag

Poesio, M. and R. Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183-216.

Recasens, M., M.A. Martí, and M. Taulé. 2007. Where anaphora and coreference meet. Annotation in the Spanish CESS-ECE corpus. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP2007)*, Borovets, Bulgaria, forthcoming.

Tutin, A., F. Trouilleux, C. Clouzot, E. Gaussier, A. Zaenen, S. Rayot, and G. Antoniadis. 2000. Annotating a large corpus with anaphoric links. In *Proceedings of the 3rd Discourse Anaphora and Anaphor Resolution Colloquium (DAARC2000)*, Lancaster.

Vieira, R. 1998. *Definite Description Processing in Unrestricted Texts.* Ph.D. Thesis, University of Edinburgh, Centre for Cognitive Science.

Vieira, R., S. Salmon-Alt, C. Gasperin, E. Schang, and G. Othero. 2002. Coreference and anaphoric relations of demonstrative noun phrases in a multilingual corpus. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC2002)*, Lisbon.

Webber, B. 1988. Discourse deixis: reference to discourse segments. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics (ACL'88)*, New York, 113-122.