

# A Chain-starting Classifier of Definite NPs in Spanish

Marta Recasens

CLiC - Centre de Llenguatge i Computació  
Department of Linguistics  
University of Barcelona  
08007 Barcelona, Spain  
mrecasens@ub.edu

## Abstract

Given the great amount of definite noun phrases that introduce an entity into the text for the first time, this paper presents a set of linguistic features that can be used to detect this type of definites in Spanish. The efficiency of the different features is tested by building a rule-based and a learning-based *chain-starting* classifier. Results suggest that the classifier, which achieves high precision at the cost of recall, can be incorporated as either a filter or an additional feature within a coreference resolution system to boost its performance.

## 1 Introduction

Although often treated together, anaphoric pronoun resolution differs from coreference resolution (van Deemter and Kibble, 2000). Whereas the former attempts to find an antecedent for each anaphoric pronoun in a discourse, the latter aims to build full coreference chains, namely linking all noun phrases (NPs) – whether pronominal or with a nominal head – that point to the same entity. The output of anaphora resolution<sup>1</sup> are noun-pronoun pairs (or pairs of a discourse segment and a pronoun in some cases), whereas the output of coreference resolution are chains containing a variety of items: pronouns, full NPs, discourse segments... Thus, coreference resolution requires a wider range of strategies in order to build the full chains of coreferent mentions.<sup>2</sup>

<sup>1</sup>A different matter is the resolution of anaphoric full NPs, i.e. those semantically dependent on a previous mention.

<sup>2</sup>We follow the ACE terminology (NIST, 2003) but instead of talking of objects in the world we talk of objects in the discourse model: we use *entity* for an object or set of objects in the discourse model, and *mention* for a reference to an entity.

One of the problems specific to coreference resolution is determining, once a mention is encountered by the system, whether it refers to an entity previously mentioned or it introduces a new entity into the text. Many algorithms (Aone and Bennett, 1996; Soon et al., 2001; Yang et al., 2003) do not address this issue specifically, but implicitly assume all mentions to be potentially coreferent and examine all possible combinations; only if the system fails to link a mention with an already existing entity, it is considered to be *chain starting*.<sup>3</sup> However, such an approach is computationally expensive and prone to errors, since natural language is populated with a huge number of entities that appear just once in the text. Even definite NPs, which are traditionally believed to refer to old entities, have been demonstrated to start a coreference chain over 50% of the times (Fraurud, 1990; Poesio and Vieira, 1998).

An alternative line of research has considered applying a filter prior to coreference resolution that classifies mentions as either chain starting or coreferent. Ng and Cardie (2002) and Poesio et al. (2005) have tested the impact of such a detector on the overall coreference resolution performance with encouraging results. Our chain-starting classifier is comparable – despite some differences<sup>4</sup> – to the detectors suggested by Ng and Cardie (2002), Uryupina (2003), and Poesio et al. (2005) for English, but not identical to strictly anaphoric ones<sup>5</sup> (Bean and Riloff, 1999; Uryupina, 2003), since a non-anaphoric NP can corefer with a previous mention.

This paper presents a corpus-based study of def-

<sup>3</sup>By *chain starting* we refer to those mentions that are the first element – and might be the only one – in a coreference chain.

<sup>4</sup>Ng and Cardie (2002) and Uryupina (2003) do not limit to definite NPs but deal with all types of NPs.

<sup>5</sup>Notice the confusing use of the term *anaphoric* in (Ng and Cardie, 2002) for describing their chain-starting filtering module.

inite NPs in Spanish that results in a set of eight features that can be used to identify chain-starting definite NPs. The heuristics are tested by building two different chain-starting classifiers for Spanish, a rule-based and a learning-based one. The evaluation gives priority to precision over recall in view of the classifier's efficiency as a filtering module.

The paper proceeds as follows. Section 2 provides a qualitative comparison with related work. The corpus study and the empirically driven set of heuristics for recognizing chain-starting definites are described in Section 3. The chain-starting classifiers are built in Section 4. Section 5 reports on the evaluation and discusses its implications. Finally, Section 6 summarizes the conclusions and outlines future work.

## 2 Related Work

Some of the corpus-driven features here presented have a precedent in earlier classifiers of this kind for English while others are our own contribution. In any case, they have been adapted and tested for Spanish for the first time.

We build a list of storage units, which is inspired by research in the field of cognitive linguistics. Bean and Riloff (1999) and Uryupina (2003) have already employed a definite probability measure in a similar way, although the way the ratio is computed is slightly different. The former use it to make a "definite-only list" by ranking those definites extracted from a corpus that were observed at least five times and never in an indefinite construction. In contrast, the latter computes four definite probabilities – which are included as features within a machine-learning classifier – from the Web in an attempt to overcome Bean and Riloff's (1999) data sparseness problem. The definite probabilities in our approach are checked with confidence intervals in order to guarantee the reliability of the results, avoiding to draw any generalization when the corpus does not contain a large enough sample.

The heuristics concerning named entities and storage-unit variants find an equivalent in the features used in Ng and Cardie's (2002) supervised classifier that represent whether the mention is a proper name (determined based on capitalization, whereas our corpus includes both weak and strong named entities) and whether a previous NP is an alias of the current mention (on the basis of a rule-based alias module that tries out different transfor-

mations). Uryupina (2003) and Vieira and Poesio (2000) also take capital and low case letters into account.

All four approaches exploit syntactic structural cues of pre- and post- modification to detect complex NPs, as they are considered to be unlikely to have been previously mentioned in the discourse. A more fine-grained distinction is made by Bean and Riloff (1999) and Vieira and Poesio (2000) to distinguish restrictive from non-restrictive post-modification by omitting those modifiers that occur between commas, which should not be classified as chain starting. The latter also list a series of "special predicates" including nouns like *fact* or *result*, and adjectives such as *first*, *best*, *only*, etc. A subset of the feature vectors used by Ng and Cardie (2002) and Uryupina (2003) is meant to code whether the NP is or not modified. In this respect, our contribution lies in adapting these ideas for the way modification occurs in Spanish – where premodifiers are rare – and in introducing a distinction between PP and AP modifiers, which we correlate in turn with the heads of simple definites.

We borrow the idea of classifying definites occurring in the first sentence as chain starting from Bean and Riloff (1999).

The precision and recall results obtained by these classifiers – tested on MUC corpora – are around the eighties, and around the seventies in the case of Vieira and Poesio (2000), who use the Penn Treebank.

Luo et al. (2004) make use of both a linking and a starting probability in their Bell tree algorithm for coreference resolution, but the starting probability happens to be the complementary of the linking one. The chain-starting classifier we build can be used to fine-tune the starting probability used in the construction of coreference chains in Luo et al.'s (2004) style.

## 3 Corpus-based Study

As fully documented by Lyons (1999), definiteness varies cross-linguistically. In contrast with English, for instance, Spanish adds the article before generic NPs (1), within some fixed phrases (2), and in postmodifiers where English makes use of bare nominal premodification (3). Altogether results in a larger number of definite NPs in Spanish and, by extension, a larger number of chain-starting definites (Recasens et al., 2009).

- (1) *Tardía incorporación de la mujer al trabajo.*  
Late incorporation of the woman to the work.  
'Late incorporation of  $\emptyset$  women into  $\emptyset$  work.'
- (2) *Villalobos dio las gracias a los militantes.*  
Villalobos gave the thanks to the militants.  
'Villalobos gave  $\emptyset$  thanks to the militants.'
- (3) *El mercado internacional del café.*  
The market international of the coffee.  
'The international  $\emptyset$  coffee market.'

Long-held claims that equate the definite article with a specific category of meaning cannot be held. The present-day definite article is a category that, although it did originally have a semantic meaning of “identifiability”, has increased its range of contexts so that it is often a grammatical rather than a semantic category (Lyons, 1999). Definite NPs cannot be considered anaphoric by default, but strategies need to be introduced in order to classify a definite as either a chain-starting or a coreferent mention. Given that the extent of grammaticization<sup>6</sup> varies from language to language, we considered it appropriate to conduct a corpus study oriented to Spanish: (i) to check the extent to which strategies used in previous work can be extended to Spanish, and (ii) to explore additional linguistic cues.

### 3.1 The corpus

The empirical data used in our corpus study come from AnCora-Es, the Spanish corpus of AnCora – Annotated Corpora for Spanish and Catalan (Taule et al., 2008), developed at the University of Barcelona and freely available from <http://clic.ub.edu/ancora>. AnCora-Es is a half-million-word multilevel corpus consisting of newspaper articles and annotated, among other levels of information, with PoS tags, syntactic constituents and functions, and named entities. A subset of 320 000 tokens (72 500 full NPs<sup>7</sup>) was used to draw linguistic features about definiteness.

### 3.2 Features

As quantitatively supported by the figures in Table 1, the split between simple (i.e. non-modified) and complex NPs seems to be linguistically relevant. We assume that the referential properties of

<sup>6</sup>*Grammaticalization*, or *grammaticalization*, is a process of linguistic change by which a content word becomes part of the grammar by losing its lexical and phonological load.

<sup>7</sup>By *full* NPs we mean NPs with a nominal head, thus omitting pronouns, NPs with an elliptical head as well as coordinated NPs.

simple NPs differ from complex ones, and this distinction is kept when designing the eight heuristics for recognizing chain-starting definites that we introduce in this section.

1. **Head match.** Ruling out those definites that match an earlier noun in the text has proved to be able to filter out a considerable number of coreferent mentions (Ng and Cardie, 2002; Poesio et al., 2005). We considered both total and partial head match, but stuck to the first as the second brought much noise. On its own, namely if definite NPs are all classified as chain starting only if no mention has previously appeared with the same lexical head, we obtain a precision (P) not less than 84.95% together with 89.68% recall (R). Our purpose was to increase P as much as possible with the minimum loss in R: it is preferred not to classify a chain-starting instance – which can still be detected by the coreference resolution module at a later stage – since a wrong label might result in a missed coreference link.

2. **Storage units.** A very grammaticized definite article accounts for the large number of definite NPs attested in Spanish (column 2 in Table 1): 46% of the total. In the light of Bybee and Hopper’s (2001) claim that language structure dynamically results from frequency and repetition, we hypothesized that specific simple definite NPs in which the article has fully grammaticized constitute what Bybee and Hopper (2001) call *storage units*: the more a specific chunk is used, the more stored and automatized it becomes. These article-noun storage units might well head a coreference chain.

With a view to providing the chain-starting classifier with a list of these article-noun storage units, we extracted from AnCora-Es all simple NPs preceded by a determiner<sup>8</sup> (columns 2 and 3 in the second row of Table 1) and ranked them by their *definite probability*, which we define as the number of simple definite NPs with respect to the number of simple determined NPs. Secondly, we set a threshold of 0.7, considering as storage units

<sup>8</sup>Only noun types occurring a minimum of ten times were included in this study. Singular and plural forms as well as masculine and feminine were kept as distinct types.

	Definite NPs	Other det. NPs	Bare NPs	Total
Simple NPs	12 739	6 642	15 183	34 564 (48%)
Complex NPs	20 447	9 545	8 068	38 060 (52%)
Total	33 186 (46%)	16 187 (22%)	23 251 (32%)	72 624 (100%)

Table 1: Overall distribution of full NPs in AnCora-Es (subset).

those definites above the threshold. In order to avoid biased probabilities due to a small number of observed examples in the corpus, a 95 percent confidence interval was computed. The final list includes 191 storage units, such as *la UE* ‘the EU’, *el euro* ‘the euro’, *los consumidores* ‘the consumers’, etc.

3. **Named entities (NEs).** A closer look at the list of storage units revealed that the higher the definite probability, the more NE-like a noun is. This led us to extrapolate that the definite article has completely grammaticized (i.e. lost its semantic load) before simple definites which are NEs (e.g. *los setenta* ‘the seventies’, *el Congreso de Estados Unidos* ‘the U.S. Congress’<sup>9</sup>), and so they are likely to be chain-starting.
4. **Storage-unit variants.** The fact that some of the extracted storage units were variants of a same entity gave us an additional cue: complementing the plain `head_match` feature by adding a gazetteer with variants (e.g. *la Unión Europea* ‘the European Union’ and *la UE* ‘the EU’) stops the `storage_unit` heuristic from classifying a simple definite as chain starting if a previous equivalent unit has appeared.
5. **First sentence.** Given that the probability for any definite NP occurring in the first sentence of a text to be chain starting is very high, since there has not been time to introduce many entities, all definites appearing in the first sentence can be classified as chain starting.
6. **AP-preference nouns.** Complex definites represent 62% out of all definite NPs (Table 1). In order to assess to what extent the referential properties of a noun on its own depend on its combinatorial potential to occur with

either a prepositional phrase (PP) or an adjectival phrase (AP), complex definites were grouped into those containing a PP (49%) and those containing an AP<sup>10</sup> (27%). Next, the probability for each noun to be modified by a PP or an AP was computed. The results made it possible to draw a distinction – and two respective lists – between PP-preference nouns (e.g. *el inicio* ‘the beginning’) and nouns that prefer an AP modifier (e.g. *las autoridades* ‘the authorities’). Given that APs are not as informative as PPs, they are more likely to modify storage units than PPs. Nouns with a preference for APs turned out to be storage units or behave similarly. Thus, simple definites headed by such nouns are unlikely to be coreferent.

7. **PP-preference nouns.** Nouns that prefer to combine with a PP are those that depend on an extra argument to become referential. This argument, however, might not appear as a nominal modifier but be recoverable from the discourse context, either explicitly or implicitly. Therefore, a simple definite headed by a PP-preference noun might be anaphoric but not necessarily a coreferent mention. Thus, grouping PP-preference nouns offers an empirical way for capturing those nouns that are *bridging* anaphors when they appear in a simple definite. For instance, it is not rare that, once a specific company has been introduced into the text, reference is made for the first time to its director simply as *el director* ‘the director’.
8. **Neuter definites.** Unlike English, the Spanish definite article is marked for grammatical gender. Nouns might be either masculine or feminine, but a third type of definite article, the neuter one (*lo*), is used to nominalize adjectives and clauses, namely “to create a referential entity” out of a non-nominal

<sup>9</sup>The underscore represents multiword expressions.

<sup>10</sup>When a noun was followed by more than one modifier, only the syntactic type of the first one was taken into account.

Given a definite mention  $m$ ,

1. If  $m$  is introduced by a neuter definite article, classify as chain starting.
2. If  $m$  appears in the first sentence of the document, classify as chain starting.
3. If  $m$  shares the same lexical head with a previous mention or is a storage-unit variant of it, classify as coreferent.
4. If the head of  $m$  is PP-preference, classify as chain starting.
5. If  $m$  is a simple definite,
  - (a) and the head of  $m$  appears in the list of storage units, classify as chain starting.
  - (b) and the head of  $m$  is AP-preference, classify as chain starting.
  - (c) and  $m$  is an NE, classify as chain starting.
  - (d) Otherwise, classify as coreferent.
6. Otherwise (i.e.  $m$  is a complex definite), classify as chain starting.

Figure 1: Rule-based algorithm.

item. Since such neuters have a low coreferential capacity, the classification of these NPs as chain starting can favour recall.

## 4 Chain-starting Classifier

In order to test the linguistic cues outlined above, we build two different chain-starting classifiers: a rule-based model and a learning-based one. Both aim to detect those definite NPs for which there is no need to look for a previous reference.

### 4.1 Rule-based approach

The first way in which the linguistic findings in Section 3.2 are tested is by building a rule-based classifier. The heuristics are combined and ordered in the most efficient way, yielding the hand-crafted algorithm shown in Figure 1. Two main principles underlie the algorithm: (i) simple definites tend to be coreferent mentions, and (ii) complex definites tend to be chain starting (if their head has not previously appeared). Accordingly, Step 5 in Figure 1 finishes by classifying simple definites as coreferent, and Step 6 complex definites as chain starting. Before these last steps, however, a series of filters are applied corresponding to the different heuristics. The performance is presented in Table 2.

## 4.2 Machine-learning approach

The second way in which the suggested linguistic cues are tested is by constructing a learning-based classifier. The Weka machine learning toolkit (Witten and Frank, 2005) is used to train a J48 decision tree on a 10-fold cross-validation. A total of eight learning features are considered: (i) head match, (ii) storage-unit variant, (iii) is a neuter definite, (iv) is first sentence, (v) is a PP-preference noun, (vi) is a storage unit, (vii) is an AP-preference noun, (viii) is an NE. All features are binary (either “yes” or “no”). We experiment with different feature vectors, incrementally adding one feature at a time. The performance is presented in Table 3.

## 5 Evaluation

A subset of AnCora-CO-Es consisting of 60 Spanish newspaper articles (23 335 tokens, 5 747 full NPs) is kept apart for the test corpus. AnCora-CO-Es is the coreferentially annotated AnCora-Es corpus, following the guidelines described in (Recasens et al., 2007). Coreference relations were annotated manually with the aid of the PALinkA (Orasan, 2003) and AnCoraPipe (Bertran et al., 2008) tools. Interestingly enough, the test corpus contains 2 575 definite NPs, out of which 1 889 are chain-starting (1401 chain-starting definite NPs are actually isolated entities), namely 73% definites head a coreference chain, which implies that a successful classifier has the potential to rule out almost three quarters of all definite mentions.

Given that chain starting is the majority class and following (Ng and Cardie, 2002), we took the “one class” classification as a naive baseline: all instances were classified as chain starting, giving a precision of 71.95% (first row in Tables 2 and 3).

### 5.1 Performance

Tables 2 and 3 show the results in terms of precision (P), recall (R), and  $F_{0.5}$ -measure ( $F_{0.5}$ ).  $F_{0.5}$ -measure,<sup>11</sup> which weights P twice as much as R, is chosen since this classifier is designed as a filter for a coreference resolution module and hence we want to make sure that the discarded cases can be really discarded. P matters more than R.

Each row incrementally adds a new heuristic to the previous ones. The score is cumulative. Notice that the order of the features in Table 2 does

<sup>11</sup> $F_{0.5}$  is computed as  $\frac{1.5PR}{0.5P+R}$ .

Cumulative Features	P (%)	R (%)	F <sub>0.5</sub> (%)
Baseline	71.95	100.0	79.37
+Head match	84.95	89.68	86.47
+Storage-unit variant	85.02	89.58	86.49
+Neuter definite	85.08	90.05	86.68
+First sentence	85.12	<b>90.32</b>	<b>86.79</b>
+PP preference	85.12	90.32	86.79
+Storage unit	89.65**	71.54**	82.67
+AP preference	<b>89.70**</b>	71.96**	82.89
+Named entity	89.20*	78.22**	85.21

Table 2: Performance of the rule-based classifier.

Cumulative Features	P (%)	R (%)	F <sub>0.5</sub> (%)
Baseline	71.95	100.0	79.37
+Head match	<b>85.00</b>	89.70	86.51
+Storage-unit variant	85.00	89.70	86.51
+Neuter definite	85.00	90.20	86.67
+First sentence	85.10	90.40	86.80
+PP preference	85.10	90.40	86.80
+Storage unit	83.80	93.50**	86.80
+AP preference	83.90	<b>93.60**</b>	<b>86.90</b>
+Named entity	83.90	93.60**	86.90

Table 3: Performance of the learning-based classifier (J48 decision tree).

not directly map the order as presented in the algorithm (Figure 1): the `head_match` heuristic and the `storage-unit_variant` need to be applied first, since the other heuristics function as filters that are effective only if head match between the mentions has been first checked. Table 3 presents the incremental performance of the learning-based classifier for the different sets of features.

Diacritics \*\* ( $p < .01$ ) and \* ( $p < .05$ ) indicate whether differences in P and R between the reduced classifier (`head_match`) and the extended ones are significant (using a one-way ANOVA followed by Tukey’s post-hoc test).

## 5.2 Discussion

Although the central role played by the `head_match` feature has been emphasized by prior work, it is striking that such a simple heuristic achieves results over 85%, raising P by 13 percentage points. All in all, these figures can only be slightly improved by some of the additional features. These features have a different effect on each approach: whereas they improve P (and decrease R) in the hand-crafted algorithm, they improve R (and decrease P) in the decision tree. In the first case, the highest R is achieved with the first four features, and the last three features

obtain an increase in P statistically significant yet accompanied by a decrease in R also statistically significant. We expected that the second block of features would favour P without such a significant drop in R.

The drop in P in the decision tree is not statistically significant as it is in the rule-based classifier. Our goal, however, was to increase P as much as possible, since false positive errors harm the performance of the subsequent coreference resolution system much more than false negative errors, which can still be detected at a later stage. The very same attributes might prove more efficient if used as additional learning features within the vector of a coreference resolution system rather than as an independent pre-classifier.

From a linguistic perspective, the fact that the linguistic heuristics increase P provides support for the hypotheses about the grammaticized definite article and the existence of storage units. We carried out an error analysis to consider those cases in which the features are misleading in terms of precision errors. The `first_sentence` feature, for instance, results in an error in (4), where the first sentence includes a coreferent NP.

- (4) La expansión de la piratería en *el Sudeste de Asia* puede destruir las economías de *la región*.  
‘The expansion of piracy in *South-East Asia* can destroy the economies of *the region*.’

Classifying PP-preference nouns as chain starting fails when a noun like *el protagonista* ‘the protagonist’, which could appear as the first mention in a film critique, happens to be previously mentioned with a different head. Likewise, not using the same head in cases such as *la competición* ‘the competition’ and *la Liga* ‘the League’ accounts for the failure of the `storage_unit` or `named_entity` feature, which classify the second mention as chain starting. On the other hand, some recall errors are due to `head_match`, which might link two NPs that despite sharing the same head point to a different entity (e.g. *el grupo Agnelli* ‘the Agnelli group’ and *el grupo industrial Montedison* ‘the industrial group Montedison’).

## 6 Conclusions and Future Work

The paper presented a corpus-driven chain-starting classifier of definite NPs for Spanish, pointing out and empirically supporting a series of linguistic features to be taken into account. Given that definiteness is very much language de-

pendent, the AnCora-Es corpus was mined to infer some linguistic hypotheses that could help in the automatic identification of chain-starting definites. The information from different linguistic levels (lexical, semantic, morphological, syntactic, and pragmatic) in a computationally not expensive way casts light on potential features helpful for resolving coreference links. Each resulting heuristic managed to improve precision although at the cost of a drop in recall. The highest improvement in precision (89.20%) with the lowest loss in recall (78.22%) translates into an  $F_{0.5}$ -measure of 85.21%. Hence, the incorporation of linguistic knowledge manages to outperform the baseline by 17 percentage points in precision.

Priority is given to precision, since we want to assure that the filter prior to coreference resolution module does not label as chain starting definite NPs that are coreferent. The classifier was thus designed to minimize false positives. No less than 73% of definite NPs in the data set are chain starting, so detecting 78% of these definites with almost 90% precision could have substantial savings. From a linguistic perspective, the improvement in precision supports the linguistic hypotheses, even if at the expense of recall. However, as this classifier is not a final but a prior module, either a filter within a rule-based system or one additional feature within a larger learning-based system, the shortage of recall can be compensated at the coreference resolution stage by considering other more sophisticated features.

The results here presented are not comparable with other existing classifiers of this type for several reasons. Our approach would perform differently for English, which has a lower number of definite NPs. Secondly, our classifier has been evaluated on a corpus much larger than prior ones such as Uryupina's (2003). Thirdly, some classifiers aim at detecting non-anaphoric NPs, which are not the same as chain-starting. Fourthly, we have empirically explored the contribution of the set of heuristics with respect to the head\_match feature. None of the existing approaches compares its final performance in relation with this simple but extremely powerful feature. Some of our heuristics do draw on previous work, but we have tuned them for Spanish and we have also contributed with new ideas, such as the use of storage units and the preference of some nouns for a specific syntactic type of modifier.

As future work, we will adapt this chain-starting classifier for Catalan, fine-tune the set of heuristics, and explore to what extent the inclusion of such a classifier improves the overall performance of a coreference resolution system for Spanish. Alternatively, we will consider using the suggested attributes as part of a larger set of learning features for coreference resolution.

## Acknowledgments

We would like to thank the three anonymous reviewers for their suggestions for improvement. This paper has been supported by the FPU Grant (AP2006-00994) from the Spanish Ministry of Education and Science, and the Lang2World (TIN2006-15265-C06-06) and Ancora-Nom (FFI2008-02691-E/FILO) projects.

## References

- Chinatsu Aone and Scott W. Bennett. 1996. Applying machine learning to anaphora resolution. In S. Wermter, E. Riloff and G. Scheler (eds.), *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Springer Verlag, Berlin, 302-314.
- David L. Bean and Ellen Riloff. 1999. Corpus-based identification of non-anaphoric noun phrases. In *Proceedings of the ACL 1999*, 373-380.
- Manuel Bertran, Oriol Borrega, Marta Recasens, and Bàrbara Soriano. 2008. AnCoraPipe: A tool for multilevel annotation. *Procesamiento del Lenguaje Natural*, 41:291-292.
- Joan Bybee and Paul Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In J. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*. John Benjamins, Amsterdam, 1-24.
- Kari Fraurud. 1990. Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7:395-433.
- Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla, and Salim Roukos. 2004. A mention-synchronous coreference resolution algorithm based on the Bell tree. In *Proceedings of ACL 2004*.
- Christopher Lyons. 1999. *Definiteness*. Cambridge University Press, Cambridge.
- Vincent Ng and Claire Cardie. 2002. Identifying anaphoric and non-anaphoric noun phrases to improve coreference resolution. In *Proceedings of COLING 2002*.
- NIST. 2003. ACE Entity detection and tracking. V.2.5.1.

- Constantin Orasan. 2003. PALinkA: A highly customisable tool for discourse annotation. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*.
- Massimo Poesio and Renata Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183-216.
- Massimo Poesio, Mijail Alexandrov-Kabadjov, Renata Vieira, Rodrigo Goulart, and Olga Uryupina. 2005. Does discourse-new detection help definite description resolution? In *Proceedings of IWCS 2005*.
- Marta Recasens, M. Antònia Martí, and Mariona Taulé. 2007. Where anaphora and coreference meet. Annotation in the Spanish CESS-ECE corpus. In *Proceedings of RANLP 2007*. Borovets, Bulgaria.
- Marta Recasens, M. Antònia Martí, and Mariona Taulé. 2009. First-mention definites: more than exceptional cases. In S. Featherston and S. Winkler (eds.), *The Fruits of Empirical Linguistics*. Volume 2. De Gruyter, Berlin.
- Wee M. Soon, Hwee T. Ng, and Daniel C. Y. Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521-544.
- Mariona Taulé, M. Antònia Martí, and Marta Recasens. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Olga Uryupina. 2003. High-precision identification of discourse-new and unique noun phrases. In *Proceedings of the ACL 2003 Student Workshop*, 80-86.
- Kees van Deemter and Rodger Kibble. 2000. Squibs and Discussions: On coreferring: coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4):629-637.
- Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539-593.
- Ian Witten and Eibe Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Xiaofeng Yang, Guodong Zhou, Jian Su, and Chew L. Tan. 2003. Coreference resolution using competition learning approach. In *Proceedings of ACL 2003*. 176-183.