Mention Detection: First Steps in the Development of a Basque Coreference Resolution System

Ander Soraluze, Olatz Arregi, Xabier Arregi, Klara Ceberio and Arantza Díaz de Ilarraza
IXA Group. University of the Basque Country
ander.soraluze@ehu.es

Abstract

This paper presents the first steps in the development of a Basque coreference resolution system. We propose a mention detector system based on a linguistic study of the nature of mentions. The system identifies mentions that are potential candidates to be part of coreference chains in Basque written texts. The mention detector is rule-based and has been implemented using finite state technology. It achieves a F-measure of 77.58% under the Exact Matching protocol and of 82.81% under Lenient Matching.

1 Introduction

Coreference resolution is the key to understanding texts, and is therefore crucial in advanced NLP applications where a higher level of comprehension of the discourse leads to better performance, such as in Information Extraction (McCarthy and Lehnert, 1995), Text Summarisation (Steinberger et al., 2007), Question Answering (Vicedo and Ferrández, 2006), Machine Translation (Peral et al., 1999), Sentiment Analysis (Nicolov et al., 2008) and Machine Reading (Poon et al., 2010).

In coreference resolution, an entity is an object or set of objects in the world, while a mention is the textual reference to an entity (Doddington et al., 2004).

It is very common to divide the coreference resolution task into two main subtasks: mention detection and resolution of references (Pradhan et al., 2011). Mention detection is concerned with identifying potential mentions of entities in the text and resolution of references involves determining which mentions refer to the same entity. Although Mention Detection has close ties to named-entity recognition (NER henceforth), it is more general and complex task than NER because besides named mentions, nominal and pronominal textual references also have to be identified (Florian et al., 2010).

Our goal is to create an accurate mention detector as the basis for developing an end-to-end coreference resolution system for Basque. We especially emphasise a linguistically sophisticated understanding of the concept of mention, which we then encode by means of regular expressions using finite state technology.

As Basque is a less resourced language, there is a considerable lack of linguistic resources, which makes it particularly challenging to develop highly accurate tools for tasks like mention detection.

This paper is structured as follows. After reviewing related work in section 2, we describe in section 3 the linguistic features related to mentions in Basque texts. Section 4 presents an overview of the system, describing the resources used in preprocessing and the rules defined for mention detection. The main experimental results are outlined in section 5 and analysed in section 6. Finally, we review the main conclusions and preview future work.

2 Related Work

Coreference resolution has received some attention in the context of overall information extrac-

Proceedings of KONVENS 2012 (Main track: oral presentations), Vienna, September 20, 2012

tion. It was first included as a specific task in the programmes of the Sixth and Seventh Message Understanding Conferences (MUC-6, 1995; MUC-7, 1998). In addition, during the period 2000–2001, the Automatic Content Extraction (ACE) program effort was devoted to Entity Detection and Tracking (EDT), which consists of finding and collecting all mentions of an entity into equivalence classes that represent references to the same entity.

However, MUC and ACE were specifically designed for shared tasks on information extraction, and thus annotation decisions regarding coreferent elements were designed to task needs at the cost of linguistic accuracy (van Deemter and Kibble, 1995; Recasens, 2010).

Recently, conferences exclusively devoted to coreference resolution have been organised. SemEval-2010 Task 1 was dedicated to coreference resolution in multiple languages (Recasens et al., 2010). One year later, in the CoNLL-2011 shared task (Pradhan et al., 2011), participants had to model unrestricted coreference in OntoNotes corpora (Pradhan et al., 2007).

To resolve coreferences, one must first detect the mentions that are going to be linked in coreference chains. As several researchers point out (Stoyanov et al., 2009; Hacioglu et al., 2005; Zhekova and Kübler, 2010), the mention detection step is crucial for the accuracy of end-to-end coreference resolution systems. Errors in mention detection propagate and reduce the level of accuracy in the performance of subsequent steps. Therefore, improving the ability of coreference resolvers to identify mentions would likely improve the state-of-the-art, as indicated by studies where the impact of mention detection for the performance of a coreference resolution system has been quantified. For example, Uryupina (2008) reports that 35% of recall errors in their coreference resolution system are caused by missing mentions and in Uryupina (2010) adds that 20% of precision errors are due to inaccurate mention detection. Similarly, in (Chang et al., 2011), a system that uses gold mentions outperforms one using predicted or system mentions by a large margin, from 15% to 18% in F1 score, while Kim et al. (2011) point out that mention detection is also essential in specialised domains like biomedicine. They further observe that using gold mentions versus system mentions can change the coreference resolution performance substantially in terms of the MUC score, from 87.32% to 49.69%.

Mention detection itself is a very challenging task since expressions can have complex syntactic and semantic structures. Considering the influence that mention detection has in coreference resolution, it is clear that this task deserves more attention.

Concerning the technology used by mention detectors, two main types can be distinguished: rule-based approaches and machine learning models.

In SemEval-2010 the majority of systems used rule-based approaches (Zhekova and Kübler, 2010; Uryupina, 2010; Attardi et al., 2010; Broscheit et al., 2010); the same was true in CoNLL-2011, where only four systems used trained models. While trained models seem to be able to balance precision and recall, and to obtain a higher F-score on the mention detection task, their recall tends to be substantially lower than that achievable by rule-based systems. The low recall has negative repercussions for the whole coreference resolution system because the system has no way to recover missed mentions. Indeed, the best-performing system in CoNLL-2011, (Lee et al., 2011), was completely rule-based.

3 Linguistic Analysis of Mentions

Although coreference is a pragmatic linguistic phenomenon highly dependent on the situational context, it shows some language-specific patterns that vary according to the features of each language. With reference to the subtask of mention detection, in this section we establish what mentions we regard as potential ones to be included in a coreference chain.

In general, we take into account noun phrases (NP), focusing on the largest span of the NP. In the case of nouns complemented by subordinate clauses and coordination, we also extract the embedded NPs of larger NPs as possible candidates for a coreference chain. We propose the following mention classification:

1. **Pronouns**: In Basque, no separate forms ex-

ist for third person pronouns versus demonstrative determiners; demonstrative determiners are used as third person pronouns (Laka, 1996). Therefore, we mark the mentions formed by demonstratives used as pronouns.

(a) *LDPko buruek Mori hautatu zuten apirilean Keizo Obuchi orduko lehen ministroa ordezkatzeko, [hark] tronbosia izan ostean.*
"The heads of LDP chose Mori in April, to replace the Prime Minister Keizo Obuchi, [who (he)] suffered a thrombosis."

2. **Possessives**: We consider two types of possessives: NPs containing a possessive determiner, even if it is not the head of that NP as in (b), and possessive pronouns as in (c).

(b) *Epitieren kasuan [[bere] helburua] lortu dezakela dirudi eta baliteke denboraldia Lehen Mailan hastea.*
"In the case of Epitie, it seems that he could achieve [[his] aim] and possibly start the football season in the Premier League."

(c) *Escuderok euskal musika tradizionala eraberritu eta indartu zuen. [Harenak] dira, esate baterako, Illeta, Pinceladas Vascas eta Eusko Salmoa obrak.*
"Escudero renewed and gave prominence to traditional Basque music. The works Illeta, Pinceladas Vascas and Eusko Salmoa, for example, are [his]"

3. **Verbal nouns**: Verbs that have been nominalised and function as the head of the mention, with the corresponding case marking suffix. The whole clause governed by the verbal noun has to be annotated.

(d) *[Instalazio militarrak ixtea] eskatuko dute.*
"They will ask for [closing the military installations]."

4. **NPs as part of complex postpositions**: Basque has a postpositional system, and therefore we mark the independent NP that goes before the complex postpositions. In (e) the postposition is *aurka* ("against"), and we annotate the noun (in that case, a proper noun) that precedes it.

(e) *Joan den astean [Moriren] aurka aurkeztutako zentsura mozioak piztu zuen LDPko krisia.*
"Last week the vote of no confidence against [Mori] caused the crisis in the LDP."

5. **NPs containing subordinate clauses**: The head of these mentions is always a noun complemented by a subordinate clause. In (f) the head noun is complemented by a subordinate clause of the type that is, for Basque, called a *complementary clause*. We take the whole stretch of the NP (both the subordinate clause and the head noun) as a mention.

In addition, relative clauses can add information to nouns as in (g). In that case the boundaries of the mention are set from the beginning of the relative clause to the end of the NP.

(f) *[DINAk Argentinan egindako krimenak ikertzeko baimena] eman du Txileko Gorte Gorenak.*
"The Supreme Court of Chile has given [permission to investigate the crimes DINA committed in Argentina]."

(g) *[Igandeko partiduak duen garrantzia] dela eta, lasai egotea beharrezkoa dutela esan zuen Lotinak.*
"Lotina said that it is necessary to stay calm because of [the importance that Sunday's match has]."

6. **Ellipsis**: In Basque the ellipsis is a broad phenomenon.

At a morphosyntactical level, a noun-ellipsis occurs when the suffixes attached to the word correspond to a noun, although the noun is not explicit in the word. We consider this type of ellipsis in the case of verbs that take suffixes indicating noun-ellipsis, as in example (i). The POS given by the analyser indicates the presence of the ellipsis phenomenon, which is implied by the presence of both the verb (*sailkatu zen-* "finished") and the ellipsis (*-Ø-ak* 'who...'). All the information corresponding to both units is stored and treated as a noun.

(i) *[Bigarren sailkatu zenak] segundo bakarra kendu zion.*
"[Ø[1] who finished in second place] only had a second's advantage."

At sentence level, the subject, object or indirect element of the sentence can be elided. The morphological information about these elements (number, person...) is given by the verb. We do not mark these elliptical pronouns as mentions (j).

(j) *Ø Ez zuen podiumean izateko itxaropen handirik.*
"[He] did not have much hope of being on the podium."

7. **Coordination**: In the case of coordination, nominal groups of a conjoined NP are extracted. We also regard as mentions the nested NPs (*siesta*, "a nap" and *atsedena* "a rest") and the whole coordinated structure (*siesta eta atsedena* "a nap and rest").

(h) *Bazkal ondoren [[siesta] eta [atsedena]] besterik ez zuten egin.*
"After lunch they did nothing but have a [[nap] and [rest]]."

Finally, we want to remark that phrases composed solely of an adjective or an adverb are not included in this annotation, because the majority of coreference relations occurs between NPs.

4 System Overview

In this section, we first introduce the preprocessing tools that our mention detector uses and then explain the rules implemented to identify the mention types described in section 3.

4.1 Preprocessing

The mention detector receives as input the results of two analysers: IXAti (Aduriz and Díaz de Ilarraza, 2003), which identifies chunks based on rule-based grammars, and ML-IXAti, which identifies clauses by combining rule-based grammars and machine learning techniques (Arrieta, 2010).

Both IXAti and ML-IXAti make use of the output produced by several tools implemented in our research group [2]:

- **Morphosyntactic analyser:** *Morpheus* (Alegria et al., 1996) performs word segmentation and PoS tagging. The module that identifies syntactic functions is implemented in the *Constraint Grammar* formalism (Karlsson et al., 1995).

- **Lemmatisation and syntactic function identifier:** *Eustagger* (Alegria et al., 2002) resolves the ambiguity caused at the previous phase.

- **Multi-words items identifier:** The aim is to determine which of two or more words are to be considered multi-word expressions (Alegria et al., 2004).

- **Named entity recogniser:** *Eihera* (Alegria et al., 2003) identifies and classifies named entities (person, organisation, location) in the text.

4.2 Defined Rules

Preprocessing by generic NLP tools, while helpful, did not by itself succeed in correctly setting the boundaries of potential mentions. To address this deficiency, we developed a mention detector consisting of a set of hand-crafted rules which have been compiled into Finite State Transducers (FST).

The use of Finite State Technology enables the processing of large datasets at a high processing speed and with low memory usage. Using *foma*[3] (Hulden, 2009), an open source platform for finite-state automata and transducers, we defined 7 FSTs, composed of 24 hand-crafted rules.

Our FSTs match NPs and clauses provided by the preprocessing tools and identify the mentions and their boundaries. They are organised into

[1] In this case Ø refers to someone.

[2] http://ixa.si.ehu.es/Ixa

[3] The regular expression syntax used in *foma* can be consulted at http://code.google.com/p/foma/

seven categories according to the linguistic analysis described in section 3.

1. **Pronouns**: Although the IXAti tool returns most pronouns, it does miss a few. This first transducer, which is the simplest one, identifies the pronouns missed in the preprocessing step.

2. **Possessives**: This FST identifies possessive pronouns and possessive determiners that are nested in another NP. In example (b) in section 3, the NP *bere helburua* "his aim" is obtained in the preprocessing step. The FST extracts from this NP a new mention, *bere* "his".

3. **Verbal nouns**: The FST identifies verbal nouns and sets the boundaries of mentions so as to group them with the linguistic elements related to them. The transducer first identifies a verbal noun in a sentence (*ixtea* "closing" in (d) in section 3) and sets the right side boundary after it. Then, the left side boundary is established at the closest clause tag proposed by the ML-IXAti tagger. It is worth mentioning that simple NPs that are nested in the verbal noun mention are marked by the syntactic analyser.

4. **Postpositional phrases**: In the case of complex postpositions a FST has been defined to modify the postpositional structure in order to obtain as a mention only the NP that is part of that structure. See (e) in section 3.

5. **NPs containing subordinate clauses**: When the FST finds a head noun of a NP complemented by a subordinate clause, it sets the right side boundary after the head (*baimena* "permission" in (f) in section 3). To set the left side boundary the transducer uses the closest boundary tag proposed by the clause tagger (*DINAk* "DINA").

   Relative clauses are treated in a manner similar to complementary ones: The mention boundaries are set from the beginning of the relative clause to the end of the NP.

6. **Ellipsis**: The FST uses the syntactic analysis to identify verbs with an elided noun. The right side boundary is established after the verb and the left side boundary is obtained using the closest clause tag.

7. **Coordination**: The identification of mentions that are components of coordination structures has been the most difficult task because it is not evident in which cases additional mentions should be obtained. This FST extracts mentions surrounding coordinating conjunctions (*eta, edo...* "and, or," etc.). It is applied only to mentions that other FSTs have identified previously and that contain a coordinating conjunction.

To better illustrate the behaviour of the mention detector we use one example to analyse the entire process of the module. Figure 1 shows the input that the system receives for the sentence *Armada britainiarrak Ipar Irlandan dituen bi kuartel eta beste bi begiratoki eraitsi dituzte.* "Two military barracks and two other viewing points that the British army has in Northern Ireland have been demolished". The figure shows the phrase tags (BNP, ENP, BVP and EVP) and clause boundaries (CB) provided by the two tools, IXAti and ML-IXAti. In addition to the NPs provided –which are themselves counted as mentions– we want to obtain from this sentence the new mention *Armada britainiarrak Ipar Irlandan dituen bi kuartel eta beste bi begiratoki*.

Based on the input received, we define the rule in Figure 2 to perform relative clause detection. First, a relative verb (RV) is defined as the composition of a verb with a relative suffix. RV is used in the definition of the relative clause mention (RM).

The rule identifies the verb containing a relative suffix (*dituen* "that has") which is tagged with VREL. Next, the right side boundary is established in the NP or coordinated NPs that follow the relative verb (*bi kuartel eta beste bi begiratoki* "Two military barracks and two other viewing points"). Finally, the left side boundary of the mention is established at the closest clause boundary ({CB) to the left (*Armada* "army"). Following these steps, the system obtains a new correct mention and tags the whole structure to show where the mention begins (<MENTION>) and ends (</MENTION>).

	[Armada	britainiarrak]	[Ipar	Irlandan]	dituen	[bi	kuartel]	eta	[beste	bi	begiratoki]	eraitsi	dituzte
IXAti	BNP	ENP	BNP	ENP	REL	BNP	ENP	PJ	BNP		ENP	BVP	EVP
ML-IXAti	{CB{CB				CB}								CB}

Figure 1: The input received by the mention detector. BNP = Begin-NP, ENP = End-NP, BVP = Begin-VP, EVP = End-VP, VREL = Relative Verb, PJ = Conjunction, CB = Clause Boundary

```
define RV Verb & $[``VREL''];
define RM [CB W+ RV NP [[and|or] NP]*] @-> "<MENTION>" ... "</MENTION>";
```

Figure 2: A simplified rule to recognise relative clauses. NP = Noun Phrase, CB = Clause Boundary, W=Word, RV = Relative Verb, RM = Relative Mention

5 Experimental setup

Many authors propose that mention identification and coreference resolution task evaluations should be carried out separately. Recasens (2010) notes that the task of mention identification is clearly distinct from coreference resolution, and that therefore, one single metric giving the overall result is less informative than evaluating each task individually. Separate evaluations allow for more detailed observation of the shortcomings in each task and thus can help determine whether a system performs well at identifying coreference links but poorly at detecting mention boundaries, or *vice versa*. In this vein, (Popescu-Belis et al., 2004) propose to consider mention identification its own task and to separate its evaluation from the evaluation of coreference resolution.

The evaluation of mention identification reveals how efficiently the system detects the mentions that are to be resolved in the coreference resolution step. Commonly used measures are precision, recall and F-measure. To calculate these, the set of manually annotated mentions (GOLD) and those extracted by the mention detector (SYS) are compared.

Typically mentions are considered correct if their span is within the span of the gold mention and contains the head word (Kummerfeld et al., 2011). This matching type is known as *Lenient Matching* or *Partial Matching*. However, stricter variations of evaluation metrics have also been applied. CoNLL-2011 Shared Task (Pradhan et al., 2011), for example, used the *Strict Matching* method, which considers only exact matches to be correct. We evaluated our mention detector using both *Lenient Matching* and *Strict Matching*.

The corpus used to develop the system is part of EPEC (the Reference Corpus for the Processing of Basque) (Aduriz et al., 2006). EPEC is composed of articles published in 2000 in *Euskaldunon Egunkaria*, a Basque language newspaper. We divided the dataset used into two main parts: one part to develop the system, with 278 mentions and the other to test it, with 384 mentions. These two parts have been manually tagged by an expert linguist.

To set the baseline, we counted as mentions the chunks (except verbal ones) proposed by the chunk identifier and compared them with gold mentions. Table 1 shows the scores obtained by our mention detector under *Exact Matching* and *Lenient Matching*, respectively, in comparison with the baseline.

	Baseline			Mention Detector		
	P	R	F_1	P	R	F_1
EM	63.37	70.33	66.65	76.85	78.59	77.58
LM	72.01	79.75	75.65	81.96	83.97	82.81

Table 1: The baseline and system scores. EM=Exact Matching, LM=Lenient Matching

6 Discussion

Using *Exact Matching* the system obtains a F-measure of 77.58%; using *Lenient Matching* the score is considerably better, 82.81%. The difference between the scoring protocols is due to *Lenient Matching* being less strict than *Exact Matching*.

We can affirm that the improvement obtained by our mention detector is significant. The system outperforms the baseline by 11 points when using *Exact Matching*, and by 7 points when the evaluation is carried out using *Lenient Matching*. We are positive that the improvement achieved in mention detection will benefit the whole process.

We carried out a qualitative evaluation to clarify why the precision and recall scores are so similar, and found that in the majority of cases the proposed mentions are exactly the same as the gold mentions. Therefore, we can argue that our system obtains high-quality mentions.

The qualitative evaluation also afforded us an opportunity to better understand the cause of errors committed by the mention detector. We observed that most of the errors are provoked by the automatic preprocessing tools. The main cause of these errors is that the span of the NPs returned by these tools exceed the gold mention's boundaries. It is obvious that these mentions will be penalised using the two scoring protocols.

Our system uses automatically processed resources; neither gold syntactic analysis nor gold NP boundaries are provided. The use of gold resources would lead to near-perfect performance in the mention detection task. Nevertheless, our main goal is to create an end-to-end coreference resolution system able to perform the entire process even when no gold resources are provided.

7 Conclusions and Future Work

We present a mention detector implemented by means of finite state transducers. The mention detector is based in a deep linguistic analysis of mentions in Basque. As many authors have affirmed, the mention detection task is crucial to the performance of a coreference resolution system. Yet the question of defining a mention is usually treated only superficially. Our hypothesis was that carrying out a detailed linguistic study of mentions and properly defining the linguistic features of mentions would produce improved results.

The scores our system achieves are very promising, especially considering that they represent the first steps in Basque coreference resolution. The scores are 77.58% under the *Exact Matching* scoring protocol and 82.81% under the *Lenient Matching* protocol.

In the future we aim to enhance the performance of the mention detector by defining a number of specific rules able to obtain mention boundaries more accurately. In addition, once IXAti and ML-IXAti are improved, our mention detector's performance will also improve substantially because the majority of errors in our system are caused by the input provided by these two tools.

We also intend to use the mention detector to automatically tag the entire EPEC corpus. The automatic tagging process will be post-edited by expert linguists and the mentions corrected, thus resulting in a gold standard corpus. We expect this corpus to be a valuable resource for taking the next steps in coreference resolution.

Acknowledgments

This work has been supported by Ander Soraluze's PhD grant from Euskara Errektoreordetza, the University of the Basque Country (UPV/EHU).

References

Aduriz, I., Aranzabe, M., Arriola, J., Atutxa, A., Diaz-De-Ilarraza, A., Ezeiza, N., Gojenola, K., Oronoz, M., Soroa, A., and Urizar, R. (2006). Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing. pages 1–15. Rodopi. Book series: Language and Computers.

Aduriz, I. and Díaz de Ilarraza, A. (2003). Morphosyntactic disambiguation ands shallow parsing in Computational Processing of Basque. *Inquiries into the lexicon-syntax relations in Basque*, pages 1–21. University of the Basque Country.

Alegria, I., Ansa, O., Artola, X., Ezeiza, N., Gojenola, K., and Urizar, R. (2004). Representation and Treatment of Multiword Expressions in Basque. In *ACL workshop on Multiword Expressions*, pages 48–55.

Alegria, I., Artola, X., Sarasola, K., and Urkia, M. (1996). Automatic morphological analysis of Basque. *Literary & Linguistic Computing*, 11(4):193–203.

Alegria, I., Ezeiza, N., Fernandez, I., and Urizar, R. (2003). Named Entity Recognition and Classification for texts in Basque. In *II Jornadas de Tratamiento y Recuperación de*

Información, (JOTRI 2003), pages 198–203, Madrid, Spain.

Alegria, I. n., Aranzabe, M., Ezeiza, N., Ezeiza, A., and Urizar, R. (2002). Using finite state technology in natural language processing of basque. In *Implementation and Application of Automata*, volume 2494 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin / Heidelberg.

Arrieta, B. (2010). *Azaleko sintaxiaren tratamendua ikasketa automatikoko tekniken bidez: euskarako kateen eta perpausen identifikazioa eta bere erabilera koma-zuzentzaile batean*. PhD thesis, University of the Basque Country.

Attardi, G., Rossi, S. D., and Simi, M. (2010). TANL-1: Coreference resolution by parse analysis and similarity clustering. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, (SemEval 2010), pages 108–111, Stroudsburg, PA, USA.

Broscheit, S., Poesio, M., Ponzetto, S. P., Rodriguez, K. J., Romano, L., Uryupina, O., Versley, Y., and Zanoli, R. (2010). BART: A multilingual anaphora resolution system. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, (SemEval 2010), pages 104–107, Stroudsburg, PA, USA.

Chang, K.-W., Samdani, R., Rozovskaya, A., Rizzolo, N., Sammons, M., and Roth, D. (2011). Inference protocols for coreference resolution. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, (CoNLL 2011), pages 40–44, Stroudsburg, PA, USA.

Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., and Weischedel, R. (2004). The Automatic Content Extraction (ACE) Program–Tasks, Data, and Evaluation. In *Proceedings of Language Resources and Evaluation Conference*, (LREC 2004), pages 837–840.

Florian, R., Pitrelli, J. F., Roukos, S., and Zitouni, I. (2010). Improving mention detection robustness to noisy input. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, (EMNLP 2010), pages 335–345, Stroudsburg, PA, USA.

Hacioglu, K., Douglas, B., and Chen, Y. (2005). Detection of entity mentions occurring in english and chinese text. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, (HLT 2005), pages 379–386, Stroudsburg, PA, USA.

Hulden, M. (2009). Foma: a finite-state compiler and library. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, (EACL 2009), pages 29–32, Stroudsburg, PA, USA.

Karlsson, F., Voutilainen, A., Heikkila, J., and Anttila, A. (1995). *Constraint Grammar, A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter.

Kim, Y., Riloff, E., and Gilbert, N. (2011). The Taming of Reconcile as a Biomedical Coreference Resolver. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 89–93, Portland, Oregon, USA.

Kummerfeld, J. K., Bansal, M., Burkett, D., and Klein, D. (2011). Mention Detection: Heuristics for the OntoNotes annotations. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, (CoNLL 2011), pages 102–106, Stroudsburg, PA, USA.

Laka, I. (1996). A Brief Grammar of Euskara, the Basque Language. http://www.ehu.es/grammar. University of the Basque Country.

Lee, H., Peirsman, Y., Chang, A., Chambers, N., Surdeanu, M., and Jurafsky, D. (2011). Stanford's multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, (CoNLL 2011), pages 28–34, Stroudsburg, PA, USA.

McCarthy, J. F. and Lehnert, W. G. (1995). Using decision trees for conference resolution. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, (IJCAI 1995), pages 1050–1055, San Francisco, CA, USA.

MUC-6 (1995). Coreference Task Definition (v2.3, 8 Sep 95). In *Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 335–344, Columbia, Maryland, USA.

MUC-7 (1998). Coreference Task Definition (v3.0, 13 Jul 97). In *Proceedings of the 7th Message Understanding Conference (MUC-7)*, Fairfax, Virginia, USA.

Nicolov, N., Salvetti, F., and Ivanova, S. (2008). Sentiment Analysis: Does Coreference Matter? In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, pages 37–40.

Peral, J., Palomar, M., and Ferrández, A. (1999). Coreference-oriented interlingual slot structure & machine translation. In *Proceedings of the Workshop on Coreference and its Applications*, (CorefApp 1999), pages 69–76, Stroudsburg, PA, USA.

Poon, H., Christensen, J., Domingos, P., Etzioni, O., Hoffmann, R., Kiddon, C., Lin, T., Ling, X., Mausam, Ritter, A., Schoenmackers, S., Soderland, S., Weld, D., Wu, F., and Zhang, C. (2010). Machine reading at the University of Washington. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, (FAM-LbR 2010), pages 87–95, Stroudsburg, PA, USA.

Popescu-Belis, A., Rigouste, L., Salmon-Alt, S., and Romary, L. (2004). Online Evaluation of Coreference Resolution. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1507–1510, Lisbon, Portugal.

Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Weischedel, R., and Xue, N. (2011). CoNLL-2011 shared task: modeling unrestricted coreference in OntoNotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, (CoNLL 2011), pages 1–27, Stroudsburg, PA, USA.

Pradhan, S. S., Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., and Weischedel, R. (2007). OntoNotes: A Unified Relational Semantic Representation. In *Proceedings of the International Conference on Semantic Computing*, (ICSC 2007), pages 517–526, Washington, DC, USA. IEEE Computer Society.

Recasens, M. (2010). *Coreference: Theory, Annotation, Resolution and Evaluation*. PhD thesis, University of Barcelona.

Recasens, M., Màrquez, L., Sapena, E., Martí, M. A., Taulé, M., Hoste, V., Poesio, M., and Versley, Y. (2010). SemEval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, (SemEval 2010), pages 1–8, Stroudsburg, PA, USA.

Steinberger, J., Poesio, M., Kabadjov, M. A., and Jeek, K. (2007). Two uses of anaphora resolution in summarization. *Information Processing and Management*, 43(6):1663–1680.

Stoyanov, V., Gilbert, N., Cardie, C., and Riloff, E. (2009). Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 656–664, Suntec, Singapore.

Uryupina, O. (2008). Error Analysis for Learning-based Coreference Resolution. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Uryupina, O. (2010). Corry: A system for coreference resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 100–103, Uppsala, Sweden. Association for Computational Linguistics.

van Deemter, K. and Kibble, R. (1995). On Coreferring: Coreference in MUC and Related Annotation Schemes. *Computational Linguistics*, 26:629–637.

Vicedo, J. and Ferrández, A. (2006). Coreference In Q&A. In *Advances in Open Domain Question Answering*, volume 32 of *Text, Speech and Language Technology*, pages 71–96. Springer.

Zhekova, D. and Kübler, S. (2010). UBIU: A language-independent system for coreference resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, (SemEval 2010), pages 96–99, Stroudsburg, PA, USA.