

Coreference Resolution for Morphologically Rich Languages. Adaptation of the Stanford System to Basque.*

*Resolución de coreferencia para lenguajes morfológicamente ricos.
Adaptación del sistema de Stanford al euskera.*

Ander Soraluze, Olatz Arregi, Xabier Arregi and Arantza Díaz de Ilarraza
Ixa Group. University of the Basque Country
{ander.soraluze, olatz.arregi, xabier.arregi, a.diazdeilarraza}@ehu.eus

Resumen: Este artículo presenta el proceso de adaptación del sistema de resolución de coreferencia de Stanford para el euskera, un idioma aglutinante, de núcleo final y *pro-drop*. Este sistema ha sido integrado en una cadena de análisis lingüística de manera que recibe como entrada textos procesados y analizados para el euskera. Hemos demostrado que haciendo uso de las características lingüísticas del lenguaje se puede mejorar la resolución de la coreferencia. En el caso de los lenguajes aglutinantes el uso de características morfosintácticas mejora claramente el rendimiento del sistema obteniéndose un incremento en CoNLL F_1 de 5 puntos para el caso de menciones automáticas y de 7,87 puntos con menciones *gold*.

Palabras clave: coreferencia, euskera, lenguaje aglutinante

Abstract: This paper presents the adaptation of the Stanford coreference resolution system to Basque, an agglutinative head-final *pro-drop* language. The adapted system has been integrated into a global linguistic analysis pipeline so that the input of the system are original Basque raw texts linguistically processed, and annotated. We demonstrate that language-specific characteristics have a noteworthy effect on coreference resolution. In the case of agglutinative languages the use of morphosyntactic features improves substantially the system’s performance, obtaining a gain in CoNLL F_1 results of 5 points when automatic mentions are used and of 7.87 points when gold mentions are provided.

Keywords: coreference, Basque, agglutinative language

1 Introduction

Coreference resolution consists of identifying textual expressions (mentions) that refer to real-world objects (entities) and determining which of these mentions refer to the same entity. It is well known that coreference resolution is helpful in NLP applications where a higher level of comprehension of the discourse leads to better performance. Information Extraction, Question Answering, Machine Translation, Sentiment Analysis, Machine Reading, Text Summarization, and Text Simplification, among others, can benefit from coreference resolution.

In this paper we present the adaptation of the Stanford Deterministic Coreference Resolution System (henceforth SDCRS) (Lee et

al., 2013) to resolve coreferences in Basque written texts. The SDCRS applies a succession of ten independent deterministic coreference models (or “sieves”) to resolve coreference in written texts. During the adaptation process, firstly, we have created a baseline system which receives as input texts processed by Basque analysis tools and uses specifically adapted static lists to identify language dependent features. Afterwards, improvements over the baseline system have been applied, adapting and replacing some of the original sieves. Our final goal is to create a robust end-to-end coreference resolution system for the Basque language.

Basque is a non-Indo-European language and differs considerably in grammar from the languages spoken in its surroundings. It is, indeed, an agglutinative head-final *pro-drop* isolated language. Furthermore, Basque is a free word order language; this means

* This work has been supported by Ander Soraluze’s PhD grant from Euskara Errektoreordetza, the University of the Basque Country (UPV/EHU) and by the Ber2Tek project, Basque Government (IE12-333).

that the order of phrases in the sentence can vary (Laka, 1996). The rich morphology of Basque requires that one takes into account the structure of words (morphological analysis) to improve coreference resolution results.

This paper is structured as follows. After reviewing related work in section 2, we describe the adaptation of the system for Basque in section 3. Section 4 presents the effects of considering some morphosyntactic characteristics of Basque and the improvements obtained with respect to the baseline system. The main experimental results are outlined in section 5 and discussed in section 6. Finally, we review the main conclusions and preview future work.

2 Related Work

The first coreference resolution systems were designed for English; nowadays, however, many conferences focus on multilingual coreference resolution. In CoNLL 2011 (Pradhan et al., 2011), participants had to model unrestricted coreference in the Ontonotes corpus (Pradhan et al., 2007) in English. Only one year later, the CoNLL 2012 shared task (Pradhan et al., 2012) involved predicting coreference in three languages, English, Chinese and Arabic. Participants adapted their systems to resolve coreference in these languages, taking into consideration the special characteristics of each language (Fernandes, dos Santos, and Milidiú, 2012). Björkelund and Farkas (2012) note that while Chinese and English are not morphologically rich languages, Arabic has a very complex morphology and this is why they had to use lemmas and unvocalized Buckwalter forms. Chen and Ng (2012), on the other hand, seek to improve the multi-pass sieve approach by incorporating lexical information using machine learning techniques. They employ different sieves depending on the language.

SemEval-2010 Task 1 (Recasens et al., 2010) was also dedicated to coreference resolution in multiple languages (Catalan, Dutch, English, German, Italian, and Spanish). This shared task addressed open questions like how much language-specific tuning is necessary to implement a general coreference resolution system portable to different languages or how helpful morphology, syntax and semantics are for solving coreference relations. Zhekova and Kübler (2010) suggest that an optimization of the feature set for

individual languages should improve system performance. Broscheit et al. (2010) affirm that substantial improvements can be achieved by incorporating language specific information with the *Language Plugin* module of their BART system. The *Language Plugin* provides an effective separation between linguistic and machine learning aspects of the problem of coreference resolution. Recently, Kopeć and Ogródniczuk (2012) have explained particularly well the process of adapting the BART system to Polish, a less-resourced language.

3 Adapting the Stanford Coreference Resolution System to Basque

The SDCRS applies a succession of ten independent deterministic coreference models (or “sieves”) one at time from highest to lowest precision. It makes use of global information through an entity-centric model that encourages the sharing of features across all mentions that point to the same real-world entity. The architecture is highly modular, which means that additional coreference models can easily be integrated. The system gives state-of-the-art performance for English and has also been incorporated into hybrid state-of-the-art coreference systems for Chinese and Arabic.

The module that resolves coreference is used at the end of a pipeline process where raw English written texts are processed. In each step, common linguistic processors (tokenizer, POS tagger, named entity recognizer) are applied to the text, thus obtaining linguistically annotated data.

The coreference resolution module is dependent on the annotations that previous modules make. As the modules were created to process English, we had to adapt the output obtained by Basque linguistic processors in order to create appropriate annotations for the coreference module.

The Basque linguistic processors used to create annotations are the following: i) A morphological analyser that performs word segmentation and PoS tagging (Alegria et al., 1996), ii) A lemmatiser that resolves the ambiguity caused at the previous phase (Alegria et al., 2002), iii) A multi-word item identifier that determines which groups of two or more words are to be considered multi-word expressions (Alegria et al., 2004), iv) A named-

entity recogniser that identifies and classifies named entities (person, organization, location) in the text (Alegria et al., 2003), v) A numerical-entity recogniser that identifies and classifies numerical entities (date, time, percent, number...) in the text (Soraluze et al., 2011), vi) A dependency parser based on Maltparser (Nivre et al., 2007); its output is then used to create constituent trees (Bengoetxea and Gojenola, 2010), and vii) A mention detector that identifies mentions that are potential candidates to be part of coreference chains in Basque written texts (Soraluze et al., 2012).

Apart from the annotations, the coreference resolution module also makes use of some static lists that are organized to exploit relevant features like gender, animacy or number. Pronouns, too, are defined as static lists. These static lists have also been adapted to Basque.

The created baseline system uses the static lists adapted to Basque and the annotations created by Basque linguistic processors. The sieves of the coreference module have not been changed at all. The results obtained by this system are presented in section 5.

4 Improvements over the baseline system

In this section we explain how we modified the baseline system taking advantage of some of the Basque language features to improve the performance of the system.

4.1 The Exact Morphology String Match sieve

Firstly, we observed the need of creating a new sieve to deal with mentions that fulfilled the string match constraint except for some grammatical suffix. This need is closely related to the agglutinative nature of Basque. The new sieve, named Exact Morphology String Match, can be considered a replacement of the original Exact String Match sieve, which links two mentions if they contain the same extent text. The Exact String Match sieve works correctly when mentions are identical. Nevertheless, this constraint is too restrictive in agglutinative languages, since the role of prepositions is played by suffixes added to word forms, for example, *lehendakariarekin* “with the president” and *lehendakariarengana* “to the president”. Ex-

act Match String sieve would not link these two mentions because their extents do not match.

In order to treat these cases correctly, the Exact Morphology Match sieve assumes that two mentions are coreferent if i) the lemmas of each word in both mentions are identical, and ii) the number and definiteness are the same. If these conditions are not fulfilled, the mentions are not considered coreferent.

The examples in Table 1 illustrate the suitability of the Exact Morphology String Match. We can see that first mention *txori politak*, and the second one, *txori politekin*, are coreferent because their lemmas are identical and they satisfy the same number and same definiteness condition. Nevertheless, although the first and third mention are identical, they are not coreferent. The first mention *Txori politak* represents a plural mention in the absolutive case, and the same string in the third row corresponds with a mention in the singular ergative case (obviously this morphological information has been previously extracted by attending to the context). Finally, the first and fourth mentions have the same lemma and number but their definiteness differs (the first is definite while the second is indefinite), so they can not be considered coreferent.

4.2 The Relaxed String Match sieve

The SDCRS has a special sieve to treat relative clauses, called Relaxed String Match. This sieve considers two mentions coreferent if their strings are identical when the text of a relative clause following the head word is dropped, e.g., “Bush” and “Bush, who was president of the U.S.”. In English relative clauses always follow the noun, but in Basque relative clauses can either follow or precede the noun. For example, the two possible equivalents in Basque for the sentence “the president, who accepted the new law” are the following: i) noun followed by the relative clause *Presidentea [zeinak lege berria onartu baitu]_{REL}* and ii) noun preceded by the relative clause *[Lege berria onartu duen]_{REL} presidentea*. Although, the two cases presented above are correct in Basque, the second one is more common.

To resolve this issue, we have adapted the Relaxed String match sieve to also take into account relative clauses that precede the

#	Mention	Translation	Lemmas	Number	Definiteness	Coreferent
1	txori politak	pretty bird	txori polit	plural	definite	-
2	txori politekin	with the pretty birds	txori polit	plural	definite	yes
3	txori politak	pretty bird	txori polit	singular	definite	no
4	txori politek	pretty birds	txori polit	plural	indefinite	no

Table 1: Examples of mentions that are coreferent and not based on their morphological features

noun. This way, the sieve is able to drop the text or relative clause preceding the head word, and mentions like *presidentea* and *lege berria onartu duen presidentea* can be linked.

4.3 The Strict Head Match sieve and its variants

The String Match sieve links two mentions if the following constraints are fulfilled: i) the candidate mention (mention to consider for resolution) head word matches any head word of mentions in the antecedent entity (*Entity Head Match*); ii) all non-stop words in the current entity to be solved are included in the set of non-stop words in the antecedent entity (*Word Inclusion*); iii) all the mention modifiers (whether nouns or adjectives) of the candidate are included in the modifiers of the antecedent (*Compatible Modifiers only*); in other words this constraint avoids clustering the mentions *autobia zuzena* “the correct motorway” and *autobia okerra* “the wrong motorway”; iv) two mentions are not in an i-within-i construct (*Not-i-within-i*), i.e, one cannot be child NP in the other’s NP constituent (Chomsky, 1981).

We have adapted the first constraint and retained the others. In our proposal, the *Entity Head Match* constraint considers for comparison the head word lemmas, number and definiteness instead of the head word forms. In this way mentions like *Vitoria-Gasteizko Eusko Legebiltzarra* “Vitoria-Gasteiz Basque Parliament” and *Vitoria-Gasteizko Legebiltzarretik* “from Vitoria-Gasteiz Parliament” that would not be clustered following the original constraint are linked by means of our new adapted sieve.

In order to improve overall F_1 by improving recall, the following three variants of the Strict Head Match sieve are applied: 1) all the constraints are considered; 2) all the constraints are considered except *Compatible Modifiers Only*; 3) all the constraints are considered except *Word Inclusion*.

As the *Entity Head Match* constraint is applied in all the variants, our adaptation influences all of them.

4.4 The Proper Head Word Match sieve

The Proper Head Word Match sieve considers two mentions coreferent if the following are fulfilled: i) the two mentions are headed by proper nouns and the head is the last word; ii) the two mentions are not in an i-within-i construct; iii) the modifiers of the two mentions cannot have location mismatches; iv) the candidate mention cannot have a number that appears in the antecedent candidate.

The first constraint of this sieve is too restrictive, because Basque is a free word order language and the last word of a mention does not obligatorily have to be the head of a mention. Therefore, we have changed the constraint in the way that the mentions headed by proper nouns can have their heads in any position within a mention. For example, the mention *Frantzia ekialdea* “eastern France” would not be a possible candidate for resolution without any changes in the sieve as the head, the proper noun *Frantzia*, is not the last word.

We also translated the list of location modifiers and the list of written numbers defined inside the sieve from English to Basque.

5 Evaluation

The corpus used to develop and test the system is a part of EPEC (the Reference Corpus for the Processing of Basque) (Aduriz et al., 2006). EPEC is a 300,000 word sample collection of news published in *Euskaldunon Egunkaria*, a Basque language newspaper. The part of the corpus we have used has about 45,000 words and it has been manually tagged by two linguists. First of all, automatically tagged mentions obtained by our mention detector have been corrected; then, coreferent mentions have been linked in clusters.

Decisions about the annotation of singletons differ depending on the corpora. In the corpus used in SemEval-2010 Task 1 (Recasens et al., 2010) all the singletons were annotated, on the contrary, in the Ontonotes

corpus (Pradhan et al., 2007) singletons were not tagged. We decided to annotate all the singletons, although they had not coreference relations in the text. In our opinion singletons are significant for a deep text understanding.

To calculate the agreement between annotators, we used the *Strict Matching* protocol which considers two mentions correct if they are identically the same. Using this protocol we compared the annotations made by the two linguists and obtained an F-measure of 94.07% for agreement. All the annotation process has been carried out using the MMAX2 annotation tool (Müller and Strube, 2006).

We divided the dataset into two main parts: one for developing the system and the other for testing. More detailed information about the two parts can be found in Table 2.

	Words	Mentions	Clusters	Singletons
Devel	30434	8432	1313	4383
Test	15949	4360	621	2445

Table 2: EPEC corpus division information

We tested two systems using the corpus: i) the baseline system (henceforth BS), which is a copy of the original SDCRS taking as input only the output of the Basque linguistic processors and translated static lists, and ii) our improved system, Basque Coreference Resolver (henceforth BCR), which modifies and adds some sieves taking advantage of the morphosyntactic features of Basque.

The metrics used to evaluate the systems are MUC (Vilain et al., 1995), B^3 (Bagga and Baldwin, 1998), $CEAF_e$ (Luo, 2005), $CEAF_m$ (Luo, 2005), and BLANC (Recasens and Hovy, 2011). The scores have been calculated using the reference implementation of the CoNLL scorer (Pradhan et al., 2014).

Table 3 shows the results obtained by each system in the test set when automatic mentions are used.

We have also evaluated the two systems using the gold mentions, i.e, when providing all the correct mentions to the coreference resolution systems. The scores obtained are shown in Table 4.

6 Discussion

In the case of automatically detected mentions, BCR outperforms the BS according to F_1 on all the metrics. In CoNLL metric, BCR

		R	P	F_1
Mention Detection		73.01	74.86	73.92
MUC	BS	22.48	35.27	27.46
	BCR	36.63	44.34	40.11
B^3	BS	54.81	66.17	59.96
	BCR	58.34	64.08	61.08
$CEAF_m$	BS	56.13	57.6	56.86
	BCR	58.52	60.00	59.25
$CEAF_e$	BS	62.08	55.5	58.61
	BCR	60.99	58.71	59.83
BLANC	BS	33.47	44.96	36.75
	BCR	39.13	47.64	42.44
CONLL	BS	-	-	48.67
	BCR	-	-	53.67

Table 3: BS and BCR scores with automatic mentions

		R	P	F_1
Mention Detection		100	100	100
MUC	BS	31.6	43.32	36.55
	BCR	51.54	56.71	54.00
B^3	BS	76.32	86.92	81.28
	BCR	81.61	86.6	84.03
$CEAF_m$	BS	72.13	72.13	72.13
	BCR	76.3	76.3	76.3
$CEAF_e$	BS	80.44	72.11	76.05
	BCR	81.00	77.97	79.46
BLANC	BS	59.47	71.06	62.94
	BCR	67.54	75.56	70.76
CONLL	BS	-	-	64.62
	BCR	-	-	72.49

Table 4: BS and BCR scores with gold mentions

has a score of 53.67, which is 5 points higher than BS, which scores 48.67.

In (Goenaga et al., 2012) an automatic coreference resolution system for Basque is presented, but unfortunately the results are not comparable with ours. The reasons of not being the works comparable are that the corpus used by the authors to evaluate pronominal anaphora resolution is not the same as the one we used, and the size of the corpora also differs considerably in size. Furthermore, some structures like relative clauses that we consider as mentions are not taken into account in the cited work.

As it is mentioned in (Pradhan et al., 2014), where official updated scores of CoNLL 2011 and CoNLL 2012 participants are presented, the best system in 2011 obtained 51.5 official score and the worst 15.5 for English. One year later in CoNLL 2012, systems obtained better results for English task: the best one scored 60.7 in CoNLL F_1 . It is worth to note that the scores from CoNLL 2011 and 2012 are not directly com-

parable with ours, given that neither the language for resolution nor the corpus used are the same.

Comparing with the results for Arabic coreference resolution, a morphologically rich language as Basque, in CoNLL 2012 the best system obtained an score of 45.2. Clearly, the results are lower for Arabic than for English. (Chen and Ng, 2012), participants in the Arabic coreference resolution task, argue that their low results for Arabic are primarily because Arabic is highly inflectional and their knowledge of the language was poor.

German is also a morphologically rich language, but while Basque is an agglutinative language German is considered fusional language. Two system presented in SemEval-2010 Task 1 (Recasens et al., 2010) that resolved coreference for German are SUCRE (Kobdani and Schütze, 2010) and UBIU (Zhekova and Kübler, 2010). SUCRE obtained an score of 55.03 CoNLL F_1 and UBIU 33.93. While SUCRE’s results are good enough, UBIU’s are quite low considering that the system is described as a language-independent for coreference resolution.

Observing our results we can affirm that the knowledge of the language, such as the morphosyntactic information, benefits considerably the coreference resolution in highly inflectional languages.

Our preliminary results are quite good in all the metrics, taking into account that there is margin of improvement as the full adaptation of the BCR is not yet finished.

When gold mentions are used our system also outperforms the baseline system according to all the metrics. The official CoNLL metric is outperformed by 7.87 points.

It is interesting to compare the effect of the mention detection. When automatic mentions are provided the CoNLL F_1 of BCR is 53.67, while providing gold mentions raises this value to 72.49. There is a considerable improvement, 18.86 points, that shows how important is to obtain a good performance in the mention detection task and how errors in this step can decrease substantially the performance on coreference resolution. Similar ideas on the importance of mention detection are presented in (Uryupina, 2008) and in (Uryupina, 2010). The scores obtained when gold mentions are provided also shows that there is still a margin to improve coreference resolution.

7 Conclusions and future work

We have adapted the SDCRS to Basque and integrated it into a global architecture of linguistic processing. Firstly, we have defined a baseline system, and afterwards, improved it based on the principle that morpho-syntactic features are crucial in the design of the sieves for agglutinative languages like Basque. The initial changes consist of the addition of a new sieve, Exact Morphology Match, replacing the original Exact String Match and the modification of Relaxed String Match sieve, Strict Head Match sieve and its variants and Proper Head Word Match sieve. Our system outperforms the baseline system in all the metrics considered. The results obtained in CoNLL metric are quite good, 53.67 when automatic mentions are used and 72.49 with gold mentions, and point that we are in a good direction to obtain a robust coreference resolution for Basque.

In the future, our aim is to analyse the influence of agglutination and free word-order on other sieves, and to implement the necessary adaptations. We also want to adapt the Pronoun Resolution sieve to Basque, taking into account the characteristics of the Basque pronouns. Nowadays, the original sieve ordering of the SDCRS is used in our system, nevertheless, better ordering options could exist. We would like to investigate which ordering of the sieves would be the optimal for Basque. It would also be interesting to carry out a deep qualitative error analysis of the results in order to obtain information about how to improve the recall of the system while preserving the precision.

We expect that the incorporation of new Basque-oriented treatments into the system will improve the overall scores.

References

- Aduriz, I., M. Aranzabe, J. M. Arriola, M. Atutxa, A. Díaz de Ilarraza, N. Ezeiza, K. Gojenola, M. Oronoz, A. Soroa, and R. Urizar. 2006. Methodology and Steps towards the Construction of EPEC, a Corpus of Written Basque Tagged at Morphological and Syntactic Levels for the Automatic Processing. Rodopi. Book series: Language and Computers., pages 1–15.
- Alegria, I., O. Ansa, X. Artola, N. Ezeiza, K. Gojenola, and R. Urizar. 2004. Representation and Treatment of Multiword

- Expressions in Basque. In *ACL workshop on Multiword Expressions*, pages 48–55.
- Alegria, I., M. Aranzabe, N. Ezeiza, A. Ezeiza, and R. Urizar. 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*. Springer Berlin / Heidelberg, pages 1–12.
- Alegria, I., X. Artola, K. Sarasola, and M. Urkia. 1996. Automatic Morphological Analysis of Basque. *Literary & Linguistic Computing*, 11(4):193–203.
- Alegria, I., N. Ezeiza, I. Fernandez, and R. Urizar. 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.
- Bagga, A. and B. Baldwin. 1998. Algorithms for Scoring Coreference Chains. In *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pages 563–566.
- Bengoetxea, K. and K. Gojenola. 2010. Application of Different Techniques to Dependency Parsing of Basque. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistically Rich Languages*, pages 31–39, Stroudsburg, PA, USA.
- Björkelund, A. and R. Farkas. 2012. Data-driven Multilingual Coreference Resolution Using Resolver Stacking. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 49–55, Jeju Island, Korea, July. Association for Computational Linguistics.
- Broscheit, S., M. Poesio, S. P. Ponzetto, K. J. Rodriguez, L. Romano, O. Uryupina, Y. Versley, and R. Zanolini. 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.
- Chen, C. and V. Ng. 2012. Combining the Best of Two Worlds: A Hybrid Approach to Multilingual Coreference Resolution. In *Joint Conference on EMNLP and CoNLL: Proceedings of the Shared Task*, pages 56–63.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Studies in generative grammar. Foris publications, Dordrecht, Cinnaminson (R.I.).
- Fernandes, E. R., C. N. dos Santos, and R. L. Milidiú. 2012. Latent Structure Perceptron with Feature Induction for Unrestricted Coreference Resolution. In *Joint Conference on EMNLP and CoNLL - Shared Task*, CoNLL '12, pages 41–48, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Goenaga, I., O. Arregi, K. Ceberio, A. Díaz de Ilaraza, and A. Jimeno. 2012. Automatic Coreference Annotation in Basque. In *11th International Workshop on Treebanks and Linguistic Theories*, Lisbon, Portugal.
- Kobdani, H. and H. Schütze. 2010. SUCRE: A Modular System for Coreference Resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, (SemEval 2010), pages 92–95, Stroudsburg, PA, USA.
- Kopeć, M. and M. Ogrodniczuk. 2012. Creating a Coreference Resolution System for Polish. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Laka, I. 1996. A Brief Grammar of Euskara, the Basque Language. <http://www.ehu.es/grammar>. University of the Basque Country.
- Lee, H., A. Chang, Y. Peirsman, N. Chambers, M. Surdeanu, and D. Jurafsky. 2013. Deterministic Coreference Resolution Based on Entity-centric, Precision-ranked Rules. *Computational Linguistics*, 39(4):885–916.
- Luo, X. 2005. On Coreference Resolution Performance Metrics. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 25–32, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Müller, C. and M. Strube. 2006. Multi-level Annotation of Linguistic Data with MMAX2. In Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Peter Lang, Frankfurt a.M., Germany, pages 197–214.
- Nivre, J., J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov, and E. Marsi. 2007. MaltParser: A language-independent System for Data-driven Dependency Parsing. *Natural Language Engineering*, 13(2):95–135.
- Pradhan, S., E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, and R. Weischedel. 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.
- Pradhan, S., X. Luo, M. Recasens, E. Hovy, V. Ng, and M. Strube. 2014. Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 30–35. Association for Computational Linguistics.
- Pradhan, S., A. Moschitti, N. Xue, O. Uryupina, and Y. Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *Proceedings of the Sixteenth Conference on Computational Natural Language Learning (CoNLL 2012)*, Jeju, Korea.
- Pradhan, S., L. Ramshaw, M. Marcus, M. Palmer, R. Weischedel, and N. Xue. 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.
- Recasens, M. and E. Hovy. 2011. BLANC: Implementing the Rand index for coreference evaluation. *Natural Language Engineering*, 17(4):485–510.
- Recasens, M., L. Màrquez, E. Sapena, M. A. Martí, M. Taulé, V. Hoste, M. Poesio, and Y. Versley. 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.
- Soraluze, A., I. Alegria, O. Ansa, O. Arregi, and X. Arregi. 2011. Recognition and Classification of Numerical Entities in Basque. In *RANLP*, pages 764–769, Hissar, Bulgaria.
- Soraluze, A., O. Arregi, X. Arregi, K. Ceborio, and A. Díaz de Ilarraza. 2012. Mention Detection: First Steps in the Development of a Basque Coreference Resolution System. In *KONVENS 2012, The 11th Conference on Natural Language Processing*, Vienna, Austria.
- 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, May. 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, July. Association for Computational Linguistics.
- Vilain, M., J. Burger, J. Aberdeen, D. Connolly, and L. Hirschman. 1995. A Model-theoretic Coreference Scoring Scheme. In *Proceedings of the 6th Conference on Message Understanding*, MUC6 '95, pages 45–52, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Zhekova, D. and S. Kübler. 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.