

Resolving Entity Coreference in Croatian with a Constrained Mention-Pair Model

Goran Glavaš and Jan Šnajder

Text Analysis and Knowledge Engineering Lab
Faculty of Electrical Engineering and Computing, University of Zagreb
Unska 3, 10000 Zagreb, Croatia
{goran.glavas, jan.snajder}@fer.hr

Abstract

Being able to identify that different mentions refer to the same entity is beneficial for applications such as question answering and text summarization. In this paper, we propose the first model for entity coreference resolution for Croatian. We enforce transitivity constraints with integer linear programming on top of pairwise decisions produced by the supervised mention-pair model. Experimental results show that the proposed model significantly outperforms two different rule-based baselines, reaching performance of 74.4% MUC score and 77.6% B^3 score.

1 Introduction

Entity coreference resolution, the task of recognizing mentions in text that refer to the same real-world entity, has been one of the central tasks of natural language processing (NLP) for decades (Grosz et al., 1983; Connolly et al., 1997; Ponzetto and Strube, 2006). Coreference resolution owes this attention to numerous applications that could greatly benefit from the ability to identify different mentions of the same entity, such as relation extraction (Shinyama and Sekine, 2006), question answering (Vicedo and Ferrández, 2000; Zheng, 2002), and text summarization (Bergler et al., 2003; Steinberger et al., 2007).

Despite being easy to define, coreference resolution is considered to be a rather difficult task, primarily because it heavily relies on external knowledge (e.g., for resolving “*U.S. President*” and “*Barack Obama*”, one needs to know that Obama is the president of the USA) (Markert et al., 2003; Durrett and Klein, 2014).

Although machine learning-based approaches to anaphora and coreference resolution for English appeared almost two decades ago (Connolly et al.,

1997), for many languages, including the majority of Slavic languages, no coreference resolution systems exist, mainly due to the lack of annotated corpora required for developing such systems.

In this paper, we present a coreference resolution model for Croatian. Our model enforces transitivity of coreference relations via integer linear programming (ILP) optimization over a set of binary coreference decisions made by the supervised mention-pair model (McCarthy and Lehnert, 1995). To the best of our knowledge, this is the first work on coreference resolution for Croatian, and one of the first efforts in coreference resolution for Slavic languages in general.

2 Related Work

Early computational approaches to coreference resolution for English were rule-based and heavily influenced by computational theories of discourse such as *focusing* and *centering* (Sidner, 1979; Grosz et al., 1983). As annotated coreference corpora became available, primarily within the Message Understanding Conferences (MUC-6 and MUC-7), research focus shifted towards supervised machine learning models. The first learning-based coreference resolution approach dates back to Connolly et al. (1997).

The mention-pair model is essentially a binary coreference classifier for pairs of entity mentions, introduced by Aone and Bennett (1995) and McCarthy and Lehnert (1995). It is still at the core of most coreference resolution systems, despite its obvious inability to enforce the transitivity inherent to the coreference relation and the fact that it requires an additional clustering algorithm to build the coreference clusters. Interestingly enough, more complex models such as entity-mention model (McCallum and Wellner, 2003; Daumé III and Marcu, 2005; Yang et al., 2008a) and ranking models (Iida et al., 2003; Yang et al., 2008b), designed to remedy for the shortcomings of the mention-pair model,

failed to demonstrate a significant performance improvements over the simple mention-pair model.

Besides for English, there is a significant body of work on coreference resolution for other major languages, including Spanish (Palomar et al., 2001; Sapena et al., 2010), Italian (Kobdani and Schütze, 2010; Poesio et al., 2010), German (Versley, 2006; Wunsch, 2010), Chinese (Converse, 2006; Kong and Zhou, 2010), Japanese (Iida et al., 2003; Iida, 2007), and Arabic (Zitouni et al., 2005; Luo and Zitouni, 2005).

On the other hand, research on coreference resolution for Slavic languages has been quite limited, mainly due to the non-existence of manually annotated corpora. The exceptions are the work done for Polish, (Marciniak, 2002; Matysiak, 2007; Kopec and Ogrodniczuk, 2012), Czech (Linh et al., 2009), and Bulgarian (Zhikov et al., 2013). In particular, Kopec and Ogrodniczuk (2012) demonstrate that a rule-based coreference resolution system for Polish significantly outperforms state-of-the-art machine learning models for English, suggesting that the coreference resolution model benefits from morphological complexity of Polish.

In this work, we present a mention-pair coreference resolution model for Croatian. Our model accounts for transitivity of coreference relations by encoding transitivity constraints as an ILP optimization problem. Our constrained mention-pair model reaches a performance of 77.6% B^3 score, which is significantly above the state-of-the-art performance for English. This supports the claim that rich morphological information facilitates coreference resolution.

3 Dataset Annotation

Supervised coreference models require a manually annotated dataset. We next describe how we compiled a coreference resolution dataset for Croatian.

3.1 Annotation Guidelines

Although coreference in most cases relates to both mentions referring to exactly the same real-world entity (i.e., identity relation), coreference may also relate to several near-identity relations between two mentions (Recasens et al., 2010); e.g., one mention may be referring to part of the entity to which the other mention refers. Arguably the most important step prior to annotating the coreference resolution dataset is to determine the identity and near-identity relations that hold between different mentions of

the same real-world entity. Considering that Croatian is a highly inflectional language, we adopt the coreference relation type scheme for inflectional languages proposed by Ogrodniczuk et al. (2013). This scheme includes the following coreference relation types (an instantiation of each of the relation types is given in Table 1):

- **IDENTITY** relation covers the most common case of coreference where both mentions refer to exactly the same real-world entity;
- **HYPER-HYPONYM** relation refers to cases where one mention is a hypernym of the other mention (but both mentions still refer to the same entity);
- **MERONYMY** relation is present where one mention refers to the part of the entity to which the other mention refers;
- **METONYMY** is a relation in which one of the mentions, although referring to the same entity as the other mention, is expressed via a phrase that typically denotes a different entity;
- **ZERO ANAPHORA** is a relation where one of the mentions is expressed implicitly in the form of a hidden subject.

Annotators were instructed to annotate instances of all of the aforementioned coreference relation types. They were instructed to link each mention to its closest previous coreferent mention in the text. Entity mentions that are not being part of at least one coreference relation were ignored.

3.2 Annotation Workflow

Six annotators participated in the annotation task. The corpus used for annotation comprised of articles from the Croatian news collection “Vjesnik”. Annotators used an in-house developed annotation tool and were provided detailed annotation guidelines. We first asked the annotators to annotate a calibration set consisting of 15 news articles. We then discussed the disagreements and resolved them by consensus.

After calibration, we conducted two rounds of annotation. In each of the rounds we paired the annotators (pairings were different between the rounds), so that we have each document annotated by exactly two annotators. In both rounds, each pair of annotators was assigned 45 news articles, but each annotator annotated the documents independently. After each of the two annotations rounds, we measured the average pairwise agreement and observed that it reached 70% of accuracy. The fol-

Coreference type	Example
IDENTITY	<i>Premijer je izjavio da on nije odobrio taj zahtjev.</i> (The Prime Minister said he didn't grant that request.)
HYPER-HYPONYM	<i>Ivan je kupio novi automobil. Taj Mercedes je čudo od auta.</i> (Ivan bought a new car . That Mercedes is an amazing car.)
MERONYMY	<i>Od jedanaestorice rukometaša danas je igralo samo njih osam.</i> (Only eight out of eleven handball players played today.)
METONYMY	<i>Dinamo Zagreb je jučer pobijedio Cibaliu. Zagrepčani su postigli tri pogotka.</i> (Dinamo Zagreb defeated Cibalia yesterday. Zagreb boys scored three goals.)
ZERO ANAPHORA	<i>Marko je išao u trgovinu. Kupio je banane.</i> (Marko went to the store. [He] bought bananas.)

Table 1: Coreference relation types.

lowing were the main causes of disagreement: (1) different pairing of mentions (80%), (2) disagreement in mention extent (16.7%), and (3) different coreference type assigned (3.3%).

The entire annotation procedure yielded a dataset consisting of 270 news articles (a total of 147,000 tokens), annotated with almost 13,000 coreference relations.¹ Expectedly, the IDENTITY relation is by far the most frequent one in the dataset, accounting for 87% of all coreference annotations, followed by MERONYMY (7%) and ZERO ANAPHORA (4%). Given the prevalence of the IDENTITY relation in our dataset, in this work we focus on extracting only coreference relations of that particular type.

4 Constrained Mention-Pair Model

At the core of our approach is a mention-pair model, i.e., a binary classifier that, given two entity mentions, predicts whether they corefer. To produce clusters of coreferent mentions, a mention-pair model needs to be coupled with two additional components: (1) a heuristic for the generation of mention-pair instances (as forming all possible pairs of mentions would result in a dataset that would be heavily skewed towards the negative class) and (2) a method for ensuring the transitivity of the coreference relation and the clustering of coreferent mentions (as the set of individual binary decisions may conflict the transitivity property of the coreference relation).

4.1 Creating Training Instances

In this work, we generate training instances using the heuristic proposed by Ng and Cardie (2002), which is, in turn, the extension of the approach by Soon et al. (2001). We thus create a positive

instance between a mention m_j and its closest preceding mention m_i , and negative instances between m_j and all the mentions in between m_i and m_j (m_{i+1}, \dots, m_{j-1}). However, if the mention m_j is non-pronominal and m_i is pronominal, then we create the positive instance by pairing m_j with its closest preceding non-pronominal mention, instead of with m_i .

4.2 Mention-Pair Model

Our mention pair model is a supervised classifier that predicts whether an IDENTITY coreference relation holds for a given pair of mentions. The classifier is based on a set of binary and numeric features, each comparing two entity mentions. Most of these features or their variants have been proposed in previous work for English and other languages. The features can be roughly grouped into four categories: string-matching features, overlap features, grammatical features, and distance-based features.

String-matching features compare the two entity mentions on the superficial string level (without any linguistic preprocessing of the mentions):

- Indication whether the two mention strings fully match (f_1);
- Indication whether one mention string contains the other (f_2);
- Length of the longest common subsequence between the mentions (f_3);
- Edit distance (i.e., Levenshtein distance) between the mentions (f_4).

Overlap features quantify the overlap between the mentions in terms of tokens these mentions share:

- Indications whether there is at least one matching word, lemma, and stem between the tokens of the two mentions (f_4 , f_5 , and f_6);
- Relative overlap between the mentions, mea-

¹A part of this dataset is freely available; cf. Section 5.

sured as the number of content lemmas (nouns, adjectives, verbs, and adverbs) found in both mentions, normalized by the token length of both mentions (f_7).

Grammatical features encode some grammatical properties and aim to indicate grammatical compatibility of the two mentions:

- Indication whether the first and second mentions are pronominal mentions, respectively (f_8 and f_9);
- Indication whether the mentions match in gender (f_{10}). Morphosyntactic descriptors for Croatian content words, including the information on gender and number, are obtained with the lemmatization tool for Croatian (Šnajder et al., 2008);
- Indication whether the mentions match in number (f_{11}).

Distance-based features indicate how far apart the two mentions are in the text (the pronominal references cannot be too far from the closest coreferent noun-phrase mention):

- Distance between the mentions in the number of tokens (f_{12});
- Distance between the mentions in the number of sentences (f_{13});
- Indication whether the two mentions are in the same sentence (f_{14});
- Indication whether the two mentions are adjacent, i.e., whether there are any other entity mentions in between them (f_{15});
- Number of other mentions in between the mentions at hand (f_{16}).

Given that our original feature space is relatively small (i.e., several orders of magnitude smaller than the number of instances in the training set), we chose as the learning algorithm the support vector machines (SVM) with the radial-basis function (RBF) kernel that maps the training instances into a high-dimensional feature space.

4.3 Enforcing Transitivity

The IDENTITY coreference relation is inherently transitive. However, by making only the local pairwise decisions, the mention-pair model does not guarantee global (i.e., document-level) coherence of its decisions with respect to the transitivity of the IDENTITY coreference relation. Thus, we need a separate mechanism to ensure that the transitivity between individual pairwise decisions holds. In

this work, we enforce transitivity as a set of linear constraints in the integer linear programming (ILP) optimization setting. We aim to maximize the objective function, which is a linear combination of mention-pair classifier confidences for individual pairwise decisions, by taking into account the linear transitivity constraints at the same time.

Objective function. Let $M = \{m_1, \dots, m_n\}$ be the set of all entity mentions in a single news article, let P be the set of all mention pairs considered by the pairwise classifier, $P = \{(m_i, m_j) \mid m_i, m_j \in M, i < j\}$, let $r(m_i, m_j)$ be the mention-pair classifier’s decision for mentions m_i and m_j , so that $r(m_i, m_j) \in \{-1, 1\}$, and let $C(m_i, m_j)$ be the confidence of the binary mention-pair classifier ($0.5 \leq C(m_i, m_j) \leq 1$). The objective function is then defined as follows:

$$\sum_{(m_i, m_j) \in P} x_{ij} \cdot r(m_i, m_j) \cdot C(m_i, m_j)$$

where x_{ij} is the binary label variable indicating whether the mentions m_i and m_j corefer.

Transitivity constraints. For all triplets of entity mentions (m_i, m_j, m_k) for which all three pairs (m_i, m_j) , (m_j, m_k) , and (m_i, m_k) exist, we enforce the following linear transitivity constraints:

$$\begin{aligned} x_{ij} + x_{jk} - x_{ik} &\leq 1, \\ x_{ij} + x_{ik} - x_{jk} &\leq 1, \\ x_{jk} + x_{ik} - x_{ij} &\leq 1, \\ \forall \{(m_i, m_j), (m_j, m_k), (m_i, m_k)\} &\subseteq P \end{aligned}$$

Clustering. After the ILP optimization, we obtain transitively coherent coreference relations, which allows us to derive the clusters of coreferent mentions simply by computing the transitive closure upon those relations.

5 Evaluation

We split the manually annotated dataset consisting of 270 documents into a train set containing 220 documents and a test set with 50 documents.² We optimized the hyperparameters of our SVM mention-pair model (C and γ) by means of 10-folded cross validation. We then trained the model with the optimal hyperparameters on the entire train set and evaluated that model on the test set.

²The test set is available from <http://takelab.fer.hr/crocoref>

Model	MUC			B^3		
	P	R	F_1	P	R	F_1
OVERLAP	81.0	42.9	54.1	75.7	54.5	61.4
GENDNUM	55.2	39.0	45.4	59.8	50.5	54.3
MP-MORPH	90.6	61.1	72.1	86.2	67.3	74.6
MP	89.4	64.7	74.2	84.0	70.1	75.4
MP+ILP	91.9	63.5	74.4	90.6	68.7	77.6

Table 2: Coreference resolution performance.

Baselines. We compare the performance of our transitively coherent mention-pair model against two different baseline models. The OVERLAP baseline classifies two mentions as coreferent if they share at least one content word. The GENDNUM baseline links each mention to the closest preceding mention with which it matches in gender and number. Standard closest-first clustering (Soon et al., 2001) is applied for both baselines.

Results. We show the performance of our mention-pair model, both without (MP) and with (MP+ILP) enforcing transitivity, along with the performance of both baselines in Table 2. We evaluate all models in terms of two standard evaluation measures for coreference resolution – MUC score and B^3 score. In order to evaluate the contribution of morphological features, we additionally evaluate the mention-pair model but excluding all features relying on morphological preprocessing (MP-MORPH).

Results show that the supervised mention-pair model significantly outperforms both reasonable rule-based baselines. When morphological features are not used, the model exhibits a slightly lower performance, although the difference is not substantial. Enforcing transitivity in an ILP setting marginally improves the overall MUC score, but yields notable 2-point improvement in B^3 score. Precision is consistently higher than recall for all models and both evaluation metrics, which is consistent with the coreference resolution results for other languages (Lee et al., 2011; Kobdani and Schütze, 2011).

Overall, our results are over 10 points higher than the state-of-the-art performance for English (Lee et al., 2011) and comparable (higher MUC and lower B^3 score) to the best results obtained for Polish (Kopec and Ogródniczuk, 2012), suggesting that coreference resolution may be easier task for morphologically complex languages.

Error analysis. In an attempt to identify the most common types of errors, we manually analyzed the errors made by the supervised mention-pair model. The vast majority of false negatives originate from mention pairs where external knowledge is necessary for inferring coreference, e.g., *željezni kancelar* (iron chancellor) and *Bismarck*. Other common causes of false negatives include abbreviations, e.g., *DS* and *Demokratski savez* (Democratic Alliance), and distant pronominal anaphora (i.e., when an anaphoric pronoun is far away from its preceding coreferent mention). Most false positives stem from non-coreferent mentions with substantial lexical overlap, e.g., *Društvo hrvatskih književnika* (Croatian Writers’ Association) and *svečanosti u Društvu hrvatskih književnika* (ceremonies at the Croatian Writers’ Association). A significant number of false positives are due to a pronominal mention being close to some non-coreferent noun-phrase mention.

6 Conclusion

We presented the first coreference resolution model for Croatian. We built a supervised mention-pair model for recognizing identity coreference relations between entity mentions and augmented it with transitivity constraints enforced via ILP optimization. We demonstrated the effectiveness of the model by showing that it substantially outperforms two rule-based baselines. Enforcing transitivity improves the B^3 score.

Manual error analysis revealed that most errors are due to the lack of external knowledge necessary for inferring coreference. Thus, we plan to extend the model with knowledge-based features obtained from external knowledge sources like Wikipedia. Furthermore, as we currently use no syntactic information, we intend to incorporate dependency relations as features.

In this work we focused on resolving identity coreference between gold event mentions. With the goal of building an end-to-end coreference resolution system for Croatian, our future efforts will focus on the development of a mention detection model. We will also consider near-identity relations like meronymy and zero anaphora.

Acknowledgments

We thank Matija Hanževački for his assistance in guiding the annotation process. We also thank the anonymous reviewers for their useful comments.

References

- Chinatsu Aone and Scott William Bennett. 1995. Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, pages 122–129.
- Sabine Bergler, René Witte, Michelle Khalife, Zhuoyan Li, and Frank Rudzicz. 2003. Using knowledge-poor coreference resolution for text summarization. In *Proceedings of the Document Understanding Conference*, pages 85–92.
- Dennis Connolly, John D Burger, and David S Day. 1997. A machine learning approach to anaphoric reference. In *New Methods in Language Processing*, pages 133–144.
- Susan Converse. 2006. *Pronominal Anaphora Resolution in Chinese*. Ph.D. thesis.
- Hal Daumé III and Daniel Marcu. 2005. A large-scale exploration of effective global features for a joint entity detection and tracking model. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 97–104.
- Greg Durrett and Dan Klein. 2014. A joint model for entity analysis: Coreference, typing, and linking. *Transactions of the Association for Computational Linguistics*, 2:477–490.
- Barbara J Grosz, Aravind K Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting on Association for Computational Linguistics*, pages 44–50.
- Ryu Iida, Kentaro Inui, Hiroya Takamura, and Yuji Matsumoto. 2003. Incorporating contextual cues in trainable models for coreference resolution. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL '03) Workshop on The Computational Treatment of Anaphora*, pages 23–30.
- Ryu Iida. 2007. *Combining Linguistic Knowledge and Machine Learning for Anaphora Resolution*. Ph.D. thesis.
- Hamidreza Kobdani and Hinrich Schütze. 2010. SUCRE: A modular system for coreference resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 92–95.
- Hamidreza Kobdani and Hinrich Schütze. 2011. Supervised coreference resolution with SUCRE. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 71–75.
- Fang Kong and Guodong Zhou. 2010. A tree kernel-based unified framework for Chinese zero anaphora resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 882–891.
- Mateusz Kopec and Maciej Ogrodniczuk. 2012. Creating a coreference resolution system for Polish. In *LREC*, pages 192–195.
- Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 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*, pages 28–34.
- Nguy Giang Linh, Václav Novák, et al. 2009. Comparison of classification and ranking approaches to pronominal anaphora resolution in Czech. In *Proceedings of the 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 276–285.
- Xiaoqiang Luo and Imed Zitouni. 2005. Multilingual coreference resolution with syntactic features. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 660–667.
- Malgorzata Marciniak. 2002. Anaphor binding for Polish. In *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium*.
- Katja Markert, Malvina Nissim, and Natalia Modjeska. 2003. Using the web for anaphora resolution. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL '03) Workshop on the Computational Treatment of Anaphora*, pages 39–46.
- Ireneusz Matysiak. 2007. Information extraction systems and nominal anaphora analysis needs. In *Proceedings of the International Multiconference on Computer Science and Information Technology*, pages 183–192.
- Andrew McCallum and Ben Wellner. 2003. Toward conditional models of identity uncertainty with application to proper noun coreference.
- Joseph F McCarthy and Wendy G Lehnert. 1995. Using decision trees for coreference resolution. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1050–1055.
- Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 104–111.
- Maciej Ogrodniczuk, Magdalena Zawisławska, Katarzyna Głowińska, and Agata Savary. 2013. Coreference annotation schema for an inflectional language. In *Computational Linguistics and Intelligent Text Processing*, pages 394–407. Springer.

- Manuel Palomar, Antonio Ferrández, Lidia Moreno, Patricio Martínez-Barco, Jesús Peral, Maximiliano Saiz-Noeda, and Rafael Muñoz. 2001. An algorithm for anaphora resolution in Spanish texts. *Computational Linguistics*, 27(4):545–567.
- Massimo Poesio, Olga Uryupina, and Yannick Versley. 2010. Creating a coreference resolution system for Italian. In *LREC*, pages 713–716.
- Simone Paolo Ponzetto and Michael Strube. 2006. Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 192–199.
- Marta Recasens, Eduard H Hovy, and Maria Antònia Martí. 2010. A typology of near-identity relations for coreference (NIDENT). In *LREC*, pages 149–156.
- Emili Sapena, Lluís Padró, and Jordi Turmo. 2010. RelaxCor: A global relaxation labeling approach to coreference resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 88–91.
- Yusuke Shinyama and Satoshi Sekine. 2006. Preemptive information extraction using unrestricted relation discovery. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 304–311.
- Candace Lee Sidner. 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Ph.D. thesis.
- Jan Šnajder, Bojana Dalbelo Bašić, and Marko Tadić. 2008. Automatic acquisition of inflectional lexica for morphological normalisation. *Information Processing & Management*, 44(5):1720–1731.
- Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational linguistics*, 27(4):521–544.
- Josef Steinberger, Massimo Poesio, Mijail Kabadjov, and Karel Ježek. 2007. Two uses of anaphora resolution in summarization. *Information Processing & Management*, 43(6):1663–1680.
- Yannick Versley. 2006. A constraint-based approach to noun phrase coreference resolution in German newspaper text. In *In Proceedings of Konferenz zur Verarbeitung Naturlicher Sprache*, pages 143–150.
- José Vicedo and Antonio Ferrández. 2000. Importance of pronominal anaphora resolution in question answering systems. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 555–562.
- Holger Wunsch. 2010. *Rule-Based and Memory-Based Pronoun Resolution for German: A Comparison and Assessment of Data Sources*. Ph.D. thesis, Universität Tübingen.
- Xiaofeng Yang, Jian Su, Jun Lang, Chew Lim Tan, Ting Liu, and Sheng Li. 2008a. An entity-mention model for coreference resolution with inductive logic programming. In *ACL*, pages 843–851.
- Xiaofeng Yang, Jian Su, and Chew Lim Tan. 2008b. A twin-candidate model for learning-based anaphora resolution. *Computational Linguistics*, 34(3):327–356.
- Zhiping Zheng. 2002. Answerbus question answering system. In *Proceedings of the Second International Conference on Human Language Technology Research*, pages 399–404.
- Valentin Zhikov, Georgi Georgiev, Kiril Simov, and Petya Osenova. 2013. Combining POS tagging, dependency parsing and coreferential resolution for Bulgarian. In *RANLP*, pages 755–762.
- Imed Zitouni, Jeff Sorensen, Xiaoqiang Luo, and Radu Florian. 2005. The impact of morphological stemming on Arabic mention detection and coreference resolution. In *Proceedings of the ACL Workshop on Computational Approaches to Semitic Languages*, pages 63–70.