

Rule-based Coreference Resolution in German Historic Novels

Markus Krug, Frank Puppe

Wuerzburg university,
Institute for Computer Science
Am Hubland
D-97074 Würzburg, Germany
markus.krug|frank.puppe@uni-wuerzburg.de

**Fotis Jannidis, Luisa Macharowsky,
Isabella Reger, Lukas Weimer**

Wuerzburg university,
Institute for German Studies
Am Hubland
D-97074 Würzburg, Germany
fotis.jannidis@uni-wuerzburg.de

Abstract

Coreference resolution (CR) is a key task in the automated analysis of characters in stories. Standard CR systems usually trained on newspaper texts have difficulties with literary texts, even with novels; a comparison with newspaper texts showed that average sentence length is greater in novels and the number of pronouns, as well as the percentage of direct speech is higher. We report promising evaluation results for a rule-based system similar to [Lee et al. 2011], but tailored to the domain which recognizes coreference chains in novels much better than CR systems like CorZu. Rule-based systems performed best on the CoNLL 2011 challenge [Pradhan et al. 2011]. Recent work in machine learning showed similar results as rule-based systems [Durrett et al. 2013]. The latter has the advantage that its explanation component facilitates a fine grained error analysis for incremental refinement of the rules.

1 Introduction

The overall goal of our research is the identification of characters in German novels from the 19th century and an analysis of their attributes. The main steps are named entity recognition (NER) of the persons, coreference resolution (CR), attribution of persons and character description with focus on sentiment analysis. While NER in novels is discussed in [Jannidis et al. 2015], we report on work in progress on rule-based coreference resolution in novels. Tests with existing rule-based or machine learning NLP tools on our novels had unsatisfying results. In contrast to newspaper texts

novels not only exhibit different topics and wording, but also show a heavy use of pronouns (in our corpus 70% of all NEs) and relative few large clusters with long coreference chains opposed to many small clusters (see baseline analysis in table 1). Another important difference is the number and lengths of passages containing direct speech [Iosif, Mishra 2014].

We decided on a rule-based approach because:

- A key aspect in coreference resolution is feature and constraint detection. Features and constraints for ruling in or out candidates for coreference with a high precision can be combined to achieve a high recall. If such features and constraints are represented by rules, the explanation component of rule-based systems is very valuable in understanding the errors and thus enabling rapid rule refinement.
- We do not have a large corpus with annotated German novels to learn from. As mentioned above, there are substantial differences between e.g. newspapers and novels, so that machine learning approaches with domain adaptation (e.g. [Yang et al. 2012]) are difficult.
- We intend to use rule-based CR to semi-automatically create a large corpus of annotated novels for experimenting with machine learning CR approaches.

We present a state-of-the-art rule-based system tailored for CR in novels. In comparison to CorZu (see section 4) which recognizes CR well in newspapers, we achieve better results in novels (MUC

F1: 85.5% vs. 65.9%, B³ F1: 56.0% vs. 33.6%). Our explanation component facilitates a fine grained error analysis for incremental rule refinement.

2 Related Work

Coreference Resolution itself is an old, but unsolved task on which a huge amount of effort was spent during the last 40 years. Large conferences like ConLL 2011 [Pradhan et al. 2011], CoNLL 2012 [Pradhan et al. 2012] and SemEval 2010 [Recasens et al. 2010]) have offered challenges for the topic not only with English text (ConLL 2011), but also for Chinese and Arabic (CoNLL 2012) and German, Dutch, Italian, Catalan and Spanish (SemEval 2010). Most approaches are based on machine learning algorithms. A large part of machine learning approaches use two phases: first classification of pairs of NEs (nouns, persons, etc.) followed by a clustering or ranking of the results (so called mention-pair model [Aone 1995, Soon et al. 2001]). Since the mention-pair model suffers from serious problems [Ng 2010], newer approaches try to match named entities directly to clusters of mentions (entity-mention model). A multitude of approaches was developed under the focus to model the affiliation of a mention to a specific entity. One goal of such an approach is to avoid problems of the mention-pair approach like ($A = B$, $B = C$, $A \neq C$, e.g. $A = \text{"Mr. Clinton"}$, $B = \text{"Clinton"}$, $C = \text{"she"}$). Aside of mention-ranking approaches [Denis, Baldrige 2008], [Rahman 2009] the system developed by Durett, Hall and Klein [Durett et al. 2013] shows that task-specific graphical models perform well following the entity-mention approach. Since rules can model hard and soft constraints directly, e.g. for pronoun resolution, rule-based approaches like the multi pass sieve for CR [Lee et al. 2011] deliver promising results, e.g. the best result in the challenge of CoNLL 2011 for English documents. Although the problem of CR has been a topic for forty years [Hobbs 1976], even good results are between 60% and 70%: [Durett et al. 2013] report a MUC-Score of 63.7% and a B³ Score of 67.8% for the CoNLL blind test set, with the system of Stanford performing comparably with 61.5% and 69.2% respectively – much worse than e.g. for NER. In [Lee et al. 2011] the Stanford system got better results with

78.6% and 80.5% showing the great influence of the data for the final results. In section 4 we therefore perform two separate experiments on two different datasets to manifest the reliability of our approach for the domain of literary novels.

3 Methods and data

Coreference resolution is based on a NLP-pipeline including tokenization, sentence splitting, part of speech tagging, lemmatization, named entity recognition and dependency parsing. In our pipeline we use the TreeTagger of Stuttgart university [Schmitt 1995] for tokenization and POS-tagging, OpenNLP¹ with the according German model for sentence splitting and the MATE-Toolkit [Bohnet 2010] for dependency parsing. Due to our overall goal, the identification and attribution of characters in German novels, we restrict coreference resolution to the resolution of persons, excluding e.g. geographic locations.

The data used for this development consists of roughly 80 segments, each sampled from a different novel. The sampling process determined a random (syntactic) sentence in the text and used all following 129 sentences, therefore forming a connected chunk of 130 sentences. This sampling process ignored the beginning of a chapter, which bears an even greater challenge for the human annotators and the algorithms, because now some segments can even start with uninformative mentions such as pronouns. With the long-term goal to get a detailed attribution of entities in the novels we developed our own annotation tool based on eclipse-rcp². Since former studies showed that the coreference task does not exhibit many ambiguities for humans our data is only annotated by one annotator.

Our corpus used in the first evaluation comprises 48 different novels. Thus, the first test corpus contains 143 000 tokens with ca. 19 000 references including proper names, personal pronouns etc., while for our second experiment we used 30 additional fragments with about 11 600 NEs and 104 000 tokens. In comparison to the German TIGER corpus [Brants et al. 2004] which consists of newspaper articles, we have on average longer sentences with 24.2 tokens compared to 16.3 to-

¹ <https://opennlp.apache.org/>

² https://wiki.eclipse.org/index.php/Rich_Client_Platform

kens. On average, one sentence contains three references to the same character or other characters. Each character appeared 10 times on average within the small novel fragments of 130 sentences, compared to ca. 4 times in the ACE2004-nwire corpus used in [Lee et al. 2011]. The majority of references are pronouns (~ 70%). In German, pronoun resolution is more ambiguous than in English, e.g. the German "sie" has three possible meanings: "she", "they", and "you" in direct speech. Only for unambiguous pronouns like "er" [he] and "ihm" [him] we can use static features like in [Lee et al. 2011]. For pronoun resolution, our rules use features of NEs like gender (male, female, neuter, unknown), number (singular, plural, unknown), person (for pronouns: first, second, third person) and whether the NE is the subject of the sentence. In general, a substantial part of a novel is direct speech, so we segment the novels in narrative and direct speech parts in a preprocessing step. In order to detect the speaker of a given direct speech annotation we use the following rules:

- Explicit speaker detection.
- Speaker propagation for longer dialogues.
- Pseudo speaker propagation for situations where in longer dialogues two persons talking in turn can be recognized, but speaker detection failed.

In order to be able to determine features like gender from non-pronoun references we use various resources like lists of male and female first names from CorZu [Klenner 2011], the morphological analysis tool of Stuttgart University SMOR [Fitschen et al. 2004], the morphological tagger from the above mentioned Mate-toolkit [Bohnet 2010] and a self-trained classifier, a maximum entropy model trained on the TIGER corpus. After applying these tools in a precision based manner (lists, own system, mate, SMOR), where the subsequent system is only used if the previous system detects "unknown", we apply a set of correction rules in order to guarantee consistency among the NEs. The heuristic rules try to infer the gender of a NE by using context clues, e.g. a subsequent reference within the same "subsentence" (that is a sentence up to the next comma) of "his" or "her" and the propagation of a recognized gender of a NE along a local chain of unambiguous NEs (e.g. for old fashioned first names like "Günderode"). Other rules exist for determination of the number of an

NE and dependency parsing is used for determining the subject of a sentence. An evaluation of the number attribute which we had annotated aside of the coreferences and NEs resulted in an accuracy of approximately 93% in the used test data. We split our documents into a small training set with just 5 documents, a first test set of 48 documents that we used to compare our performance with the system CorZu, and into another test set consisting of 30 completely unseen documents where we evaluated the robustness of our system.

Our system has a similar rule organization as [Lee et al. 2011] with passes, i.e. rule sets, which build on the results of former passes. While [Lee et al. 2011] uses 7 passes, we extend this by using 11 passes:

- 1. Pass: exact match:** All identical non-pronouns are marked as coreferent. They are also considered as coreferent if their cases differ ("Annette" vs "ANNETTE").
- 2. Pass: Nameflexion:** We designed a distance metric that allows us to detect derivations (or nicknames) from names and mark them as coreferent. ("Lydia", "Lyden", "Lydchen").
- 3. Pass: Attributes:** We use all modifiers (derived from the output of the parser) and match them against the strings of the NEs of the other cluster. Coreference occurs if there is an "equals-ignore-case"-match. ("Die alte Getrud" ,... "die Alte") ["the elderly Gertrud", "the elderly"]
- 4. Pass: precise constructs:** Appositions, relative and reflexive pronouns are assigned to the preceding NE. In addition, these pronouns get the gender and number of the NE in order to support subsequent resolution of other pronouns.
- 5-7. Pass: 5. strict head match, 6. relaxed head match and 7. title match:** These 3 passes recognize coreferent NEs, where an NE consists of several words. The first rule, named strict head match, removes all titles from the given mentions and then compares the remaining words. Two NEs are said to be coreferent if there is at least one word that appears in both mentions and they agree in number and gender ("Baron Landsfeld" , "Herr Landsfeld") ["Baron Landsfeld", "Mister Landsfeld"]. The relaxed head match only requires that one word of one NE is contained in a word of the other NE. Since titles were removed in the previous two

rules we added another rule specifically for titles and match those to the most recent NE which contains the given title.

8. Semantic pass: For this semantic pass we use the synonyms in the German resource GermaNet³, again, an agreement in gender is required. This matches for example “Gatte” and “Gemahl” [“spouse”, “consort”].

9. Pass: pronoun resolution: pronouns are resolved to the most recent, suitable precedent NE. To respect salience we sorted our previous NEs in a manner that preferred the last subject of earlier sentences over the closest NEs in those sentences. A suitable precedent is a one that doesn’t conflict with a given constraint. In the current implementation we respect the following constraints:

- Compliance in gender/number and person.
- Compliance in its context (are they both part of a direct speech or both part of a narrative segment).
- The candidate and the actual pronoun do not harm a given constraint of binding theory.

10. Pass: Detection of the addressed person in direct speech: For each direct speech annotation we try to find the addressed NE. We do this by using several handcrafted lexico-syntactic patterns (matching against expressions such as “Alexander, you are great”). Based on the results of speaker detection we use a propagation of the addressed persons in dialogues.

11. Pass: pronouns in direct speech: We then resolve all instances of <I> to the speaker and all instances of <you> to the person the speaker talks to (if known). If the speaker of two subsequent direct speech annotations doesn’t change, but the addressed person differs, we assume that the speaker only uses a different naming for the person he is talking to and therefore set these NEs as coreferent.

Fig. 1 shows the explanation, annotation and error analysis component of our rule-based tool.

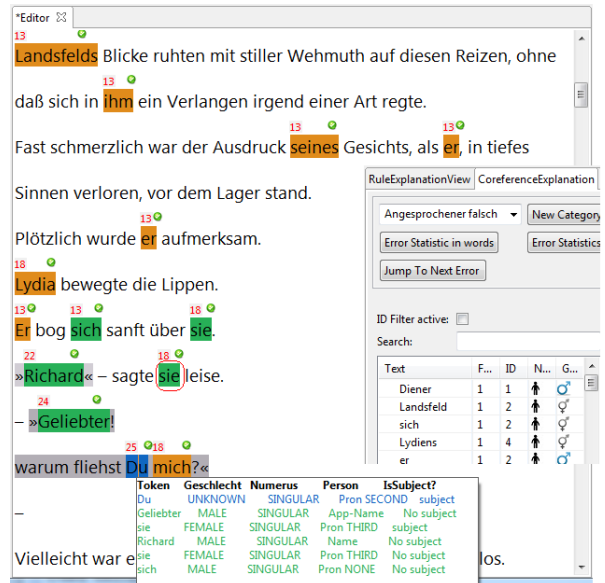


Fig. 1: Explanation and annotation editor for rule-based coreference resolution. All named entities (NE) are already marked. Identical numbers above the NE denote that they are coreferent to the same entity. For explanation, the user clicks on one NE (here: “Du” [you]) and the system shows in a pop up menu the last five NE tokens with some key attributes: Geschlecht ([gender]: male, female, neuter, unknown), number (singular, plural, unknown), Person (for pronouns: first, second third persons and for nouns the type of NE like real name, appellative name or pseudo-person) and whether the NE is the subject of the sentence. The annotator can classify errors like the missing coreference of “DU” to “Geliebter” [lover] with categories (here: “angesprochener falsch” [wrong reference]) with a drop down menu in the right frame. If there is direct speech, it is highlighted with a background color and the speaker (here “sie” [she]) is marked with a red circle.

4 Evaluation and error analysis

We evaluated the coreference resolution algorithm in two experiments. The first one uses the test corpus of 48 novel fragments with about 19.000 manually annotated character references in total. The following common evaluation metrics are used (see [Luo 2005]):

- The MUC-Score. It is based on the idea to count the minimum amount of links which need to be added to derive the true set of entities from the predicted set of entities or vice versa, divided by the amount of links in the spanning tree of the true partition. The MUC-Score itself is the harmonic mean out of both numbers that you get

³ <http://www.sfs.uni-tuebingen.de/GermaNet/>

when you switch the true partition with the gold partition.

- The B³-Score. The MUC Score cannot measure the influence of singleton clusters, that's why an additional evaluation metric is needed. The B³-Score scales the overlap of predicted clusters and true clusters, based on how many markables were correctly assigned to a given cluster.

The effect of the different evaluation measures on a newspaper corpus with rather short coreference chains and on a novel corpus with long chains is shown in the baseline analysis in table 1. While for newspapers the baseline with n clusters for n NEs is very good, for novels the baseline with just one cluster for all NEs performs well. This can be explained using the structure of the underlying entities. While in newspaper texts many different entities with only a few mentions appear, our domain shows relatively few entities that tend to show up frequently.

Baseline	corpus	MUC			B ³		
		prec.	recall	F1	prec.	recall	F1
1 coreference cluster for all n named entities	newspaper	24%	100%	39%	100%	2%	5%
	novels	89%	100%	94%	100%	21%	34%
n coreference clusters for n named entities	newspaper	0%	0%	0%	76%	100%	86%
	novels	0%	0%	0%	11%	100%	19%

Table 1: Baseline analysis for a typical newspaper and novel corpus with assigning all n NEs to either just one cluster or to n different clusters.

We compared our system with the free coreference resolution software CorZu, using ParZu⁴ [Sennrich 2009] as its parser from the university of Zurich, which was developed using a newspaper corpus. CorZu was given the same annotated named entities in the same novel fragments, so that the detected chains were comparable. Table 2 shows the results. Our system is about 20 percent points better than CorZu for both evaluation scores MUC F1 and B³ F1.

Scores in %	MUC precision	MUC recall	MUC F1	B ³ prec.	B ³ recall	B ³ F1
our system	89.1	83.2	85.5	70.5	83.2	56.0
CorZu	77.0	57.7	65.9	69.5	22.7	33.6

Table 2: evaluation results of our system and CorZu on 48 novel fragments with about 19 000 named entities.

The effect of the passes (see section 3) in our system evaluated on the novel corpus is given in table

3. For reference, we added the results from Lee et al. [Lee et al. 2011], the results of the system of Stanford, evaluated on an ACE newspaper corpus. It shows that pronoun resolution is much more important in novels than in newspapers, while exact string matches and head matches already result in rather high scores on the ACE newspaper corpus.

Scores in %	our system evaluated with the novel corpus		Stanfords Sieve evaluated with ACE newspaper corpus	
	MUC F1	B ³ F1	MUC F1	B ³ F1
Passes				
1	27.5	24.6	47.8	69.4
1-4	37.7	28.1	59.9	73.3
1-8	38.9	28.9	67.1	76.9
1-9	83.3	52.6	78.6	80.5
1-11	85.5	56.0		

Table 3: Evaluation and comparison of the effects of the different passes of the rule-based algorithm.

We finally evaluated our system on our second test set, comprising 30 completely unseen fragments and achieved an F1-score of 86% MUC-F1 and a B³-F1 of 55.5%. It is almost identical to the result of the first test set. Rule-based systems with an explanation component allow a fine-grained error analysis. Table 4 shows an error analysis for 5 randomly selected novel fragments from the 30 novels, drawn from the second test set that we used for evaluation:

Document (novel fragment)	# NES	# Cluster (Gold)	# Cluster found	MUC F1	B ³ F1	# Errors	# Wrong g n p	# Ds related	# Heuristics	# Semantic
1	332	16	64	86%	46%	58	3	22	21	12
2	185	8	22	94%	80%	16	1	4	3	8
3	261	31	44	90%	80%	22	6	0	6	10
4	283	28	47	77%	38%	48	5	2	20	21
5	469	39	39	90%	66%	61	13	1	22	25
Sum						205	28	29	72	76
Average	306	24	43	87%	62%	100%	14%	14%	35%	37%

Table 4: Number of named entities, clusters, evaluation metrics and error types for a sample of 5 novel fragments, drawn randomly from our second test set comprising 30 fragments. The category "Wrong g|n|p" refers to the sum of mistakes the algorithm made that were caused by a wrong assignment of gender, number or person. The category "Ds related", contains all errors related to direct speech, e.g. by assigning a wrong speaker or the wrong detection of the addressed person to a given direct speech annotation.

Table 4 shows that even though we combined 4 different morphological resources the recognition

⁴ http://www.cl.uzh.ch/research/coreferenceresolution_en.html

of wrong number, gender and person still makes up a fraction of about 14% of the total amount of errors in the analyzed documents. Another part with 14% of the mistakes is the category that describes all errors related to direct speech, e.g. wrong speaker detection, missed detection of “Sie” [you] in the role of “du” or wrong detection of an addressed person. We intend to find some additional constraints to further reduce the errors made in these categories. The next category with 35% error contribution, labeled as heuristics, sums up all the errors which happened due to a wrong assumption of salience, parser errors or errors that were induced by former misclassified NEs. Still the biggest share of mistakes (37%) and probably also the ones that are most difficult to fix is the class of semantic errors. Most of these misclassifications can only be resolved with additional knowledge about the world or the entities in the novel itself (“a widow is a woman who lost her husband; “his profession is forester”; ...). Apart from these, there are other mistakes related to an unmodeled context, such as thoughts, songs or letters that appear throughout the text. We plan to integrate the work of Brunner [Brunner 2015] to detect those instances and thereby improve the quality of our system.

5 Conclusion

CR for NE can be viewed as a task in which candidates for coreference are filtered out by constraints until only one candidate remains. Rule-based knowledge representation is well suited for this task. The more constraints can be modeled, the better the performance of the system. Our error analysis shows that we can cut our current error rate by roughly 28% with more precise grammatical constraints (14% for errors related to direct speech and 14% for gender, number and person related errors). However, we also plan the integration of semantic constraints and information, similar to Haghighi and Klein [Haghighi, Klein 2009]. A promising way is to collect information about the persons in the text, which is also the next step in our overall goal, the automated character analysis: determining all attributes assigned in a novel to a character.

References

- Aone, C. and Bennett, S. 1995. *Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies*. Proceedings of the 33rd annual meeting on Association for Computational Linguistics (ACL '95). Association for Computational Linguistics, 122-129.
- Bohnet, B. 2010. *Very High Accuracy and Fast Dependency Parsing is not a Contradiction*. The 23rd Int. Conf. on Computational Linguistics (COLING 2010), Beijing, China.
- Brants, S., Dipper, S., Eisenberg, P., Hansen, S., König, E., Lezius, W., Rohrer, C., Smith, G. and Uszkoreit, H. 2004. *TIGER: Linguistic Interpretation of a German Corpus*. Journal of Language and Computation 2 (4), 597-620.
- Brunner, A. 2015. *Automatische Erkennung von Redewiedergabe: Ein Beitrag zur Quantitativen Narratologie (Narratologia)*. [Automatic Recognition of Recorded Speech: a Contribution to Quantitative Narratologia] Walter De Gruyter Inc.
- Durrett G., Hall D., and Klein D. 2013. *Decentralized Entity-Level Modeling for Coreference Resolution*. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1), 114-124.
- Ferrucci, D. and Lally, A. 2004. *UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment*. Nat. Lang. Eng. 10, 327-348. <http://dx.doi.org/10.1017/S1351324904003523>
- Fitschen, A., Schmid, H. and Heid, U. 2004. *SMOR: A German computational morphology covering derivation, composition, and inflection*. Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC 2004), 1263-1266.
- Haghighi, A. and Klein, D. 2009. *Simple Coreference Resolution with Rich Syntactic and Semantic Features*. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Singapore, 1152-1161.
- Hinrichs, E., Kübler, S. and Naumann, K. 2005. *A unified representation for morphological, syntactic, semantic, and referential annotations*. Proceedings of the ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky, pages 13–20, Ann Arbor, MI.
- Hobbs, J. 1976. *Pronoun Resolution*. Technical report, Dept. of Computer Science, CUNY, Technical Report TR761.
- Iosif, E. and Mishra, T. 2014. *From Speaker Identification to Affective Analysis: A Multi-Step System for Analyzing Children's Stories*. Proceedings of the 3rd Workshop on Computational Linguistics for Literature. Gothenburg, Sweden, 40-49.

- Jannidis, F., Krug, M., Reger, I. Toepfer, M. Weimer, L., Puppe, F. 2015. *Automatische Erkennung von Figuren in deutschsprachigen Romanen*. [Automatic recognition of Characters in German novels] Digital Humanities im deutschsprachigen Raum (Dhd 2015), Graz, Austria, 2015.
- Klenner, M; Tuggener, D. 2011. *An Incremental Entity-mention Model for Coreference Resolution with Restrictive Antecedent Accessibility*. Recent Advances in Natural Language Processing (RANLP 2011), Hissar, Bulgaria, 178-185.
- 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*. Proc. of the 15th Conference on Computational Natural Language Learning: Shared Task (CONLL Shared Task '11). Association for Computational Linguistics, 28-34.
- Luo, X. 2005. *On Coreference Resolution Performance Metrics*. Proc. of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05). Association for Computational Linguistics, 25-32.
- Ng, V. 2010. *Supervised Noun Phrase Coreference Research: The First Fifteen Years*. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL '10). Association for Computational Linguistics, 1396-1411.
- Pascal, D. and Baldridge, J. 2008. *Specialized models and ranking for coreference resolution*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '08). Association for Computational Linguistics, Stroudsburg, PA, USA, 660-669.
- Pradhan, S., Moschitti, A., Xue, N., Uryupina, O. and Zhang, Y. 2012. *CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes*. In Proceedings of EMNLP and CoNLL-2012: Shared Task, 1-40.
- Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Weischedel, R., and Xue, N. 2011. *CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes*. Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task (CONLL Shared Task '11). Association for Computational Linguistics, 1-27.
- Rahman, A and Ng, V. 2009. *Supervised Models for Coreference Resolution*. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 968-977.
- Recasens, M., Martí, T., Taulé, M., Màrquez, L., and Sapena, E. 2009. *Coreference Resolution in Multiple Languages*. Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009), 70-75.
- Schmid, H. 1995. *Improvements in Part-of-Speech Tagging with an Application to German*. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland.
- Sennrich, R., Schneider, G., Volk, M., Warin, M. 2009. *A New Hybrid Dependency Parser for German*. Proceedings of GSCL Conference, Potsdam.
- Soon, W, Ng, H. and Lim, D. 2001. *A Machine Learning Approach to Coreference Resolution of Noun Phrases*. Comput. Linguist. 27 (4), 521-544.
- Yang, J., Mao, Q., Xiang, Q., Tsang, I., Chai, K., Chieu, H. 2012. *Domain Adaptation for Coreference Resolution: An Adaptive Ensemble Approach*. Proceedings of the 2012 Joint Conference on Empirical Methods in NLP and CNLL, 744-753.