

# A Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology

Ulrich Schäfer Christian Spurk Jörg Steffen

German Research Center for Artificial Intelligence (DFKI), Language Technology Lab  
Campus D 3 1, D-66123 Saarbrücken, Germany

{ulrich.schaefer,christian.spurk,joerg.steffen}@dfki.de

## ABSTRACT

We describe a large coreference annotation task performed on a corpus of 266 papers from the ACL Anthology, a publicly, electronically available collection of scientific papers in the domain of computational linguistics and language technology. The annotation comprises mainly noun phrase coreference of the full textual content of each paper in the Anthology subset. It has been performed carefully and at least twice for each paper (initial annotation and secondary correction phase). The purpose of this paper is to summarize the comprehensive annotation schema and release the corpus publicly, along with this paper. The corpus is by far larger than the ACE coreference corpora. It can be used to train coreference resolution systems in the Computational Linguistics and Language Technology domain for semantic search, taxonomy extraction, question answering, citation analysis, scientific discourse analysis, etc.

---

KEYWORDS: Coreference Resolution, Resource, Annotated Corpus, Scientific Papers, eScience.

---

## 1 Introduction and Motivation

Coreference resolution (CR), mostly on newspaper text, has been studied *in extenso* during the last decades. The task consists in finding all mentions of real-world entities such as persons or organizations in a text, regardless of their textual representation. It could be a proper name, a role (e.g. ‘the director’), a pronoun, or a similar circumscribing expression (mention).

Related (with an overlap) to this task is anaphora resolution. Here, the interpretation of a mention, called anaphor (e.g. a pronoun), depends on a previous mention, called antecedent, or on context. Different relations may hold between an anaphor and its antecedent, e.g. part-of, subset, etc. If both refer to the same entity, the anaphoric relation is also coreferential. In contrast to coreference, the order matters, but in general, not every anaphoric relation is coreferential and vice versa. A more detailed discussion of the coreference vs. anaphora distinction can be found in van Deemter and Kibble (2000).

For 15 years, significant progress in terms of robustness and coverage has been made by applying machine learning and including semantic information. Instead of enumerating the history of CR literature, we refer the interested reader to Ng (2010) and Mitkov (1999). They present comprehensive, though compact, surveys of the CR research for this period. However, most of the research so far was done in the news text domain only, as the largest available annotated corpora such as MUC (Grishman and Sundheim, 1996) and ACE (Doddington et al., 2004) mostly were of this origin.

The purpose of our endeavor is to move to a different domain, namely scientific texts, extracted

from proceedings papers and journal articles. We do this because our evaluations have shown that systems trained on news text, where mainly persons and organizations are the subject of coreference phenomena, perform worse on scientific text.

We observe a general rising research interest in applying CR to real-world texts other than newswire, e.g. the CoNLL Shared Task on Modeling Unrestricted Coreference in OntoNotes (Pradhan et al., 2011) and the BioNLP Shared Task supporting task: Protein/Gene Coreference Task (Nguyen et al., 2011).

Watson et al. (2003) argue that CR in scientific text may be harder than in newspaper text, as scientific text tends to be more complex and contains relatively high proportions of definite descriptions, which are the most challenging to resolve. Conversely, the proportion of easier-to-determine entities such as person names, is inferior in papers.

In scientific texts, anaphoric expression referring to named entities are less frequent. Instead, references to domain terminology and abstract entities such as results, variables and values are more important.

Motivation for improving CR in scientific text is manifold. Applications such as Question Answering (Watson et al., 2003), Information Extraction (Gaizauskas and Humphreys, 2000), ontology extraction, or accurate semantic search in digital libraries (Schäfer et al., 2011) can benefit from identified coreferences in running text.

On the one hand, making implicit or hidden relations explicit and complete by resolving e.g. anaphora may increase redundancy and hence the chance for answer candidates to be found. On the other hand, coreferences, if resolved correctly, may help to increase precision. Without anaphora replaced by their antecedents, certain propositions simply cannot be found. The same holds for variants of corefering expressions. A recent work also shows the benefits of CR for biomedical event extraction from scientific literature (Miwa et al., 2012).

The paper is structured as follows. In Section 2, we discuss recent related work. We present details of the coreference annotation task and parts of the annotation guidelines in Section 3. Section 4 discusses error analysis, inter-annotator agreement and correction. Finally, we give a conclusion.

## 2 Related Work

While a considerable part of previous and current research concentrates on news texts as provided by the MUC and ACE corpora, coreference annotation of scientific papers is a relatively new area. In particular, work on coreference phenomena in scientific text mostly seems to focus on the biomedical domain. There is only a small number of publications dealing with other science domains.

One of the earliest approaches is performed in the context of the Genia project and corpus of medical texts from MEDLINE. In a first stage, only MEDLINE abstracts are used (Yang et al., 2004), later *other*-anaphora, a very specific sub-task, are investigated using full paper content (Chen et al., 2008).

Gasperin (2009) presents a full annotation of anaphora and coreference in biomedical text, but only noun phrases referring to biomedical entities are considered. On the basis of this annotation, she implements a probabilistic anaphora resolution system.

In contrast, Cohen et al. (2010) build a corpus of 97 full-text journal articles in the biomedical

domain where every co-referring noun phrase is annotated (CRAFT – Colorado Richly Annotated Full Text). Their annotation guidelines follow those of the OntoNotes project (Hovy et al., 2006), adapted to the biomedical domain. OntoNotes itself is a text corpus of approx. one million words from mainly news texts (newswire, magazines, broadcast conversations, web pages). It also contains general anaphoric coreference annotations (Pradhan et al., 2007): events and (like in our annotation) unlimited noun phrase entity types.

Kim and Webber (2006) investigate a special aspect, citation sentences where a pronoun such as “they” refers to a previous citation. The study is performed on astronomy journal articles and a maximum-entropy classifier is trained.

Kaplan et al. (2009) investigate coreferences and citations as well, but only at a very small scale (4 articles from the *Computational Linguistics* journal). They focus on so-called c-sites which are the sentences following a citation that also refer to the same paper (typically by anaphora). The authors train a specific coreference model for this phenomenon. They show that exploitation of coreference chains improves the extraction of citation contexts which they then use for research paper summarization.

### 3 Corpus Creation

Our corpus comprises 266 scientific papers from the ACL Anthology (Bird et al., 2008) sections P08 (ACL-2008 long papers), D07 (EMNLP-CoNLL-2007) and C02 (COLING 2002). Texts were extracted from PDF using a commercial OCR program which guarantees uniform, though not perfect, quality of the resulting text files. We did not rely on the original ACL-ARC text files converted with PDFBox because they contained considerable extraction errors depending on the (various) PDF tools that were used to generate the PDFs. Hence quality of extraction would have depended on the PDF generator, and OCR-based extraction is much more independent from the generation process.

Moreover, PDFBox cannot reliably recover reading order from text typeset in multiple columns (again, depending on the PDF generator used). OCR introduces sporadic character and layout recognition errors, but overall works robustly, cf. discussion and a recent and even more accurate approach in Schäfer et al. (2012). The main part of the corpus creation endeavor consisted in manual annotation assisted by a customized version of the MMAX2 annotation tool (Müller and Strube, 2006), operating on the extracted raw texts (the annotators had the possibility to open and view the original PDF files). In a second step, the corpus was then augmented with automatically created annotations.

#### 3.1 Manual Annotations

The annotators were not trained in the Computational Linguistics domain, but as advanced students of English language and literature studies, they had some prior knowledge of general linguistics.

We gave our human annotators the same task that a coreference resolution system shall solve. This task is similar to the core annotation task of the ACE program: in the so-called “entity detection and tracking (EDT) task, all mentions of an entity, whether a name, a description, or a pronoun, are to be found and collected into equivalence classes based on reference to the same entity” (Dodgington et al., 2004, p. 837). However, unlike in ACE, we did not restrict the type of entities to be annotated. Because of this we do not even distinguish entity types in our annotation scheme.

In terms of ACE entity classes, we only consider referential mentions for annotation, i.e. only “Specific Referential (SPC)”, “Generic Referential (GEN)”, and “Under-specified Referential (USP)” (LDC, 2004, p. 17f.) entities, but we do not explicitly differentiate between these classes in our corpus.

Only noun phrases (NPs, including coordinations of NPs), possessive determiners (“my”, “your”, ...) and proper names (which may be part of NPs as in “Sheffield’s GATE system”) were considered as possible entity mentions. As for mention types we asked our annotators to classify each mention as one of the types listed in Table 1. It also contains the different kinds of pronouns and other noun phrases that can be mentions.

Just like the annotators in ACE, our annotators were advised to identify the maximal extent of entity mentions. The mention extent thus includes all modifiers of the mention’s head, i.e. any preceding adjectives and determiners, post-modifying phrases like prepositional phrases and relative clauses or parenthetical modifiers.

Coreferences between mention pairs were marked, i.e. coreferential entity mentions were put into the same coreference set. Only those entity mentions were marked which are coreferential with any other mention in the text, i.e. a coreference set always contains at least two mentions.

The annotators were asked to annotate *maximal coreference sets*, i.e., whenever they found two coreferential markables, they only created a new coreference set if there was not any other coreference set whose elements refer to the same referent as the two newly found markables. In other words, for every pair of document and real-world referent there should be only one coreference set with markables of the document referring to the referent. In the extreme case, coreferential markables in the abstract at the beginning of a paper and in the conclusion at the end, and all markables referring to the same entity in between were to be put into the same set. This is useful because subsets for smaller ranges (e.g. for paragraphs or sections) can easily be derived from the complete annotation.

Our manually annotated corpus contains 1,326,147 tokens (ACE 2004: 189,620) in 48,960 sentences (ACE: 5,654). The number of coreferring mentions in non-singleton coreference sets is 65,293 (ACE: 22,293). This number is plausible because scientific text typically contains less person and organization mentions than newspaper text.

Mention Type	Amount
def-np (definite NPs)	32,547
ppr (personal pronouns)	5,921
ne (proper names incl. citations)	14,451
ppos (possessive pronouns/determiners)	3,407
indef-np (indefinite NPs)	6,820
conj-np (coordinations)	1,446
pds (demonstrative pronouns)	435
prefl (reflexive pronouns)	266
$\Sigma$	65,293

Table 1: Annotated mention types and their frequency in the manually annotated corpus. Only mentions appearing in non-singleton coreference sets are counted.

## 3.2 Annotation Guidelines and Corpus Data

The full annotation guidelines (approx. 20 printed pages) with many examples and special hints for the corpus-specific phenomena such as citations, as well as user guides for the annotation tool MMAX2 are too comprehensive to be discussed here in detail. Therefore, they are part of the attached, compressed archive in file `AnnotationGuidelines.html` (along with A4 and letter paper versions in PDF format for printing). Independently of the conference proceedings, the data will be made available on `http://take.dfk1.de/#2012` and in the ACL Anthology as supplementary material to the electronic version of this publication.

The archive also contains the complete annotated data in MMAX2 format for each of the 266 papers in the subdirectory `annotation`. The file `README.txt` contains instructions on how to download, install and run MMAX2 and open the annotated corpus files for inspection. MMAX2 needs to be downloaded separately<sup>1</sup>, a screenshot is depicted in Figure 1.

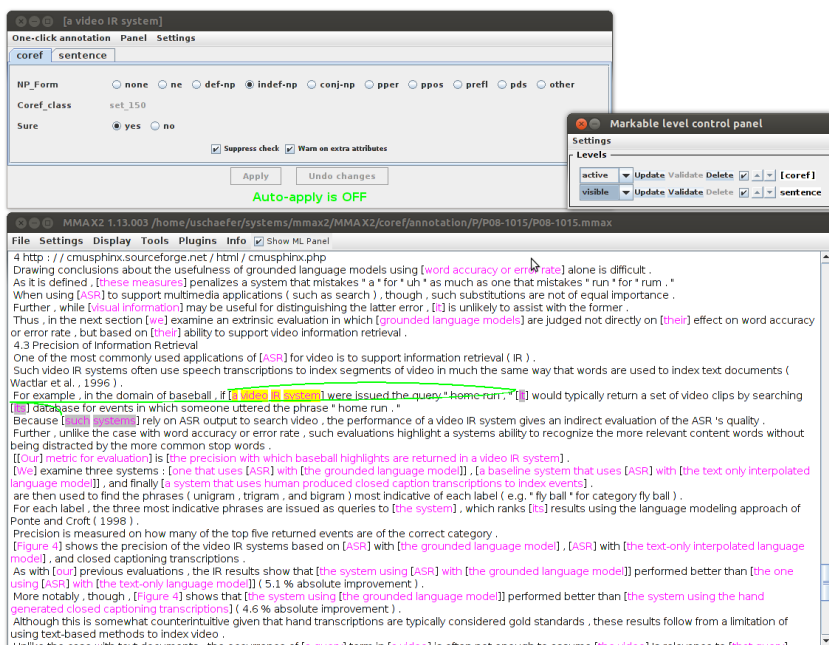


Figure 1: ACL Anthology paper annotation in the MMAX2 user interface.

The annotation guidelines introduce terminology, explain which markables to annotate and which not, both with many examples, summarize general annotation principles, and present an introduction to the MMAX2 annotation tool.

<sup>1</sup><http://sourceforge.net/projects/mmax2/files/>

### 3.2.1 Markables to Annotate

**Named Entities.** Names and named entities (NEs) are usually (definite) NPs and as such can enter into coreference relations, i.e., they are relevant markables. NEs may be, among others, names of companies, organizations, persons, locations, languages, currencies, programming languages, standards, scientific fields, systems, frameworks, etc. As a special case for our ACL Anthology corpus, we consider citations in scientific papers as NEs, too.

**Definite Noun Phrases.** Definite noun phrases are NPs which correspond to a specific and identifiable entity in a given context. In many cases this definiteness is marked by the definite article “the” or a demonstrative determiner such as “these” or “that”.

**Indefinite Noun Phrases.** Indefinite noun phrases are NPs which do not correspond to a specific and identifiable entity in a given context. In many cases this indefiniteness is marked by the indefinite articles “a” and “an” or it is indicated by the lack of a certain determiner.

**Conjunctions.** For our annotation task, we define a conjunction to be an NP which results by conjoining other NPs. The most common junctor which is used for conjoining NPs is “and”. Other junctors include, for example, “or”, “as well as” or the discontinuous junctor “both ... and”.

**Personal Pronouns.** Personal pronouns are pronouns which stand for other NPs and which are even complete NPs themselves. The most common personal pronouns in English are “I”, “you”, “he”, “she”, “they”, “it”, “me”, “him”, “us”, “them”, “her” and “we”.

**Possessive Pronouns.** Possessive pronouns in a strict sense are NPs which stand for another NP and which attribute ownership to the NP they substitute, e.g., “mine”, “hers” or “ours”. In our annotation task, we assume a broader sense in which possessive determiners are also considered to be possessive pronouns, e.g., “his”, “her” or “my”.

**Reflexive Pronouns.** Reflexive pronouns are pronouns that substitute the NP to which they refer in the same clause as the NP. The most common reflexive pronouns in English are “myself”, “yourself”, “himself”, “herself”, “themselves”, “itself”, “ourselves”, “yourselves” and “themselves”.

**Demonstrative Pronouns.** In our annotation task, demonstrative pronouns are pronouns which are NPs that stand for some other NP of the discourse. As such they are very similar to personal pronouns. The most common demonstrative pronouns in English are “this”, “that”, “these” and “those” while for the annotations in our corpus, “these” will mostly be found as markable.

**Relative Pronouns.** Relative pronouns are pronouns which introduce relative clauses. We are only interested in relative pronouns that introduce non-restrictive relative clauses. In restrictive relative clauses, the relative pronoun does not refer to any real-world entity and therefore it can't be a coreferential markable (The relative pronoun refers syntactically to the head noun of the noun phrase (NP) to which the relative clause belongs, however, this is no coreference; the NP is semantically incomplete without the relative clause). In non-restrictive clauses, the relative pronoun really corefers with the noun phrase (NP) to which the relative clause belongs. The relative pronouns in English which are also relevant for annotation are “who”, “which”, “whose”, “where”, “whom” and “when” as well as sometimes “that” and rarely “why”.

### 3.2.2 Markables not to Annotate

Our detailed annotation guidelines do not only give definitions, explanations and examples of markables and coreference phenomena to annotate, they also explain what shouldn't be annotated. Here is an excerpt.

**Relative Clauses.** In general, relative clauses alone are not markables themselves, but only part of other markables.

**Restrictive relative pronouns and clauses** should not be annotated, as the NP the relative pronoun refers to is semantically incomplete without the relative clause.

**Only Definite Predicate Nominatives with Definite Subjects.** A predicate nominative can only be a coreferential markable if it is definite and connected to a definite subject.

**Predicate nominatives** are only to be annotated if they are definite and connected with a definite subject (“A mason is a workman” is indefinite and thus not coreferential).

**Bound anaphora** (e.g. in “Every teacher likes his job.”) should not be annotated because the referents do not necessarily refer to the same.

**Indirect anaphora or bridging references** (e.g. in “The bar is crowded. The waitress is stressed out.”) should not be annotated because the referents are not identical.

### 3.3 Automatic Annotations

Because the main purpose of the corpus will be training machine learning systems, we also need examples of mentions that are not coreferential with any other mention in the text – as kind of negative examples. These single mentions were automatically annotated. To achieve this, we looked up all NPs (including coordinations functioning as NPs), possessive determiners and proper names (including citations as a special case) in the corpus that were not part of any entity mention, yet. All these were then automatically classified into the above-mentioned mention types and stored with the manually annotated mentions.

The automatic annotations were generated using the Stanford Parser (Klein and Manning, 2003) in version 1.6.3 for NPs and possessive determiners. Proper nouns are detected using the SProUT system (Drożdżyński et al., 2004) with its generic named entity grammar for English. SProUT robustly recognizes *inter alia* person names and locations in MUC style in running text, without any domain-specific adaptations or extensions except for citations. For the detection of citations, we have created an elaborate regular expression that reliably matches all kinds of citation patterns.

### 3.4 Citations

Coreferences in citation context have special properties (Kim and Webber, 2006). When they exist, they are in most cases anaphoric pairs, typically the antecedent is a name. Roughly 10 % of the sentences in the corpus contain citations. Therefore, special care has been taken to coreference phenomena in conjunction with citations.

In the ACL Anthology, citations could be quite reliably identified automatically by regular expression patterns, as the citation styles are restricted. The annotators then only had to connect with e.g. pronouns in follow-up sentences.

## 4 Error Analysis, Inter-annotator Agreement and Correction

In the initial annotation phase, 13 % of our corpus was annotated twice by different annotators in order to measure inter-annotator agreement. We did this measurement as it was done for MUC (Hirschman et al., 1997) and in the same way as a coreference resolution system is evaluated against some gold standard: one annotation was set to be the gold standard (“key”) and the second annotation was set to be the “response”. Herewith we reached an inter-annotator agreement of 49.5 MUC points (for MUC score calculation see Vilain et al. (1995)). Although the MUC measure is questionable (Luo, 2005) and the task is difficult, this number is too low and asked for improvements.

Therefore in a second phase, the annotation guidelines have been improved in order to cover more corner cases and to resolve possible ambiguities. Additionally, all annotations were checked and corrected at least a second time in order to find accidental annotation mistakes and to be consistent with the updated guidelines. This procedure has also been suggested by Hirschman et al. (1997) for the MUC data, who – after optimizing their annotation guidelines and changing the annotation process to a two-step process – could improve their inter-annotator agreement by about 12 %.

The second round has been performed by a single person over approx. 9 months part-time (8–10 hrs/week). Therefore we did not measure the inter-annotator agreement a second time, but on the other hand a single corrector ensures the annotation is of uniform quality throughout the whole corpus.

### Conclusion

We have developed a comprehensive annotation schema and annotation guidelines for coreference in scientific text and fully annotated a 266 paper subset of the ACL Anthology. The corpus is publicly available along with this paper. By a coreference resolution system built on top of it, e.g. training available tools such as LBJ (Bengtson and Roth, 2008; Rizzolo and Roth, 2010), Reconcile (Stoyanov et al., 2010, 2011), Stanford’s dcoref (Raghunathan et al., 2010; Lee et al., 2011), or (Haghighi and Klein, 2009, 2010), it could serve to improve other NLP tasks such as semantic search, taxonomy extraction, question answering, citation analysis, scientific discourse analysis, etc.

The corpus in its current state is not perfect. It will probably be necessary to add a further round of annotation assessment and correction. At this point, our project ends and we release the annotation data to the public along with the hope that the scientific community finds it useful and further improves the corpus.

### Acknowledgments

We would like to thank the student annotators, most notably Leonie Grön and Philipp Schu, for their intelligent, careful and patient work. We also thank the three anonymous reviewers for helpful comments. The work described in this paper has been funded by the German Federal Ministry of Education and Research, projects TAKE (FKZ 01IW08003) and Deependance (FKZ 01IW11003), and under the Seventh Framework Programme of the European Commission through the T4ME contract (grant agreement no.: 249119).



## References

- Bengtson, E. and Roth, D. (2008). Understanding the value of features for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP-2008*, pages 294–303, Morristown, NJ, USA.
- Bird, S., Dale, R., Dorr, B., Gibson, B., Joseph, M., Kan, M.-Y., Lee, D., Powley, B., Radev, D., and Tan, Y. F. (2008). The ACL anthology reference corpus: A reference dataset for bibliographic research. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008)*, Marrakesh, Morocco.
- Chen, B., Yang, X., Su, J., and Tan, C. L. (2008). Other-anaphora resolution in biomedical texts with automatically mined patterns. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-2008)*, pages 121–128, Manchester, UK.
- Cohen, K. B., Lanfranchi, A., Corvey, W., Jr., W. A. B., Roeder, C., Ogren, P. V., Palmer, M., and Hunter, L. (2010). Annotation of all coreference in biomedical text: Guideline selection and adaptation. In *Proceedings of the 2nd Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTextM-2010)*.
- 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 the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pages 837–840.
- Drożdżyński, W., Krieger, H.-U., Piskorski, J., Schäfer, U., and Xu, F. (2004). Shallow processing with unification and typed feature structures – foundations and applications. *Künstliche Intelligenz*, 2004(1):17–23.
- Gaizauskas, R. and Humphreys, K. (2000). Quantitative evaluation of coreference algorithms in an information extraction system. In Botley, S. and McEnery, T., editors, *Corpus-based and Computational Approaches to Discourse Anaphora*, pages 145–169. John Benjamins, Amsterdam.
- Gasparin, C. V. (2009). *Statistical anaphora resolution in biomedical texts*. PhD thesis, University of Cambridge, Cambridge, UK.
- Grishman, R. and Sundheim, B. (1996). Message understanding conference - 6: A brief history. In *Proceedings of COLING-96*, pages 466–471.
- 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*, pages 1152–1161, Singapore.
- Haghighi, A. and Klein, D. (2010). Coreference resolution in a modular, entity-centered model. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 385–393, Los Angeles, California.
- Hirschman, L., Robinson, P., Burger, J., and Vilain, M. (1997). Automating coreference: The role of annotated training data. In *Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*.
- Hovy, E., Marcus, M., Palmer, M., Ramshaw, L., and Weischedel, R. (2006). OntoNotes: The 90% solution. In *Proceedings of the HLT-NAACL 2006 Companion Volume*.

Kaplan, D., Iida, R., and Tokunaga, T. (2009). Automatic extraction of citation contexts for research paper summarization: A coreference-chain based approach. In *Workshop on text and citation analysis for scholarly digital libraries (NLP4DL), ACL-IJCNLP-09*, pages 88–95, Singapore.

Kim, Y. and Webber, B. (2006). Implicit references to citations: A study of astronomy papers. In *Proceedings of the 20th International CODATA Conference: Scientific Data and Knowledge within the Information Society*.

Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Meeting of the Association for Computational Linguistics (ACL-2003)*, pages 423–430. Association for Computational Linguistics.

LDC (2004). Annotation guidelines for entity detection and tracking (EDT) – version 4.2.6.

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 *15th CoNLL: Shared Task*, pages 28–34.

Luo, X. (2005). On coreference resolution performance metrics. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 25–32, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Mitkov, R. (1999). Anaphora resolution: The state of the art. Technical report, School of Languages and European Studies, University of Wolverhampton.

Miwa, M., Thompson, P., and Ananiadou, S. (2012). Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*.

Müller, C. and Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. In Braun, S., Kohn, K., and Mukherjee, J., editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany.

Ng, V. (2010). Supervised noun phrase coreference research: The first fifteen years. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411, Uppsala, Sweden.

Nguyen, N., Kim, J.-D., and Tsujii, J. (2011). Overview of BioNLP 2011 protein coreference shared task. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 74–82, Portland, Oregon, USA.

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

Pradhan, S. S., Ramshaw, L., Ralph, Weischedel, MacBride, J., and Micciulla, L. (2007). Unrestricted coreference: Identifying entities and events in OntoNotes. In *Proceedings of the International Conference on Semantic Computing*, pages 446–453. IEEE.

- Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C. (2010). A multi-pass sieve for coreference resolution. In *Proceedings of EMNLP-2010*, pages 492–501, Cambridge, MA.
- Rizzolo, N. and Roth, D. (2010). Learning based Java for rapid development of NLP systems. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC-2010)*, Valletta, Malta.
- Schäfer, U., Kiefer, B., Spurk, C., Steffen, J., and Wang, R. (2011). The ACL Anthology Searchbench. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 7–13, Portland, Oregon. Association for Computational Linguistics.
- Schäfer, U., Read, J., and Oepen, S. (2012). Towards an ACL Anthology Corpus with logical document structure. An overview of the ACL 2012 contributed task. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, pages 88–97, Jeju Island, Korea. Association for Computational Linguistics.
- Stoyanov, V., Babbar, U., Gupta, P., and Cardie, C. (2011). Reconciling OntoNotes: Unrestricted coreference resolution in OntoNotes with Reconcile. In *Proceedings of 15th CoNLL: Shared Task*, pages 122–126.
- Stoyanov, V., Cardie, C., Gilbert, N., Riloff, E., Buttlar, D., and Hysom, D. (2010). Coreference resolution with Reconcile. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 156–161, Uppsala, Sweden.
- van Deemter, K. and Kibble, R. (2000). On coreferring: coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4).
- Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the Fourth Message Understanding Conference (MUC-4)*, San Mateo, CA. Morgan Kaufmann.
- Watson, R., Preiss, J., and Briscoe, E. J. (2003). Contribution of domain-independent robust pronominal anaphora resolution to open-domain question answering. In *Proceedings of the International Symposium on Reference Resolution*.
- Yang, X., Su, J., Zhou, G., and Tan, C. L. (2004). An NP-cluster based approach to coreference resolution. In *Proceedings of COLING-2004*, pages 226–232, Geneva, Switzerland.

