Language Resources and Annotation Tools for Cross-Sentence Relation Extraction

Sebastian Krause, Hong Li, Feiyu Xu, Hans Uszkoreit, Robert Hummel, Luise Spielhagen
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI)
Alt-Moabit 91c, 10559 Berlin, Germany
{skrause, lihong, feiyu, uszkoreit, rohu01, masp04}@dfki.de

Abstract

In this paper, we present a novel combination of two types of language resources dedicated to the detection of relevant relations (RE) such as events or facts across sentence boundaries. One of the two resources is the *sar-graph*, which aggregates for each target relation ten thousands of linguistic patterns of semantically associated relations that signal instances of the target relation (Uszkoreit and Xu, 2013). These have been learned from the Web by intra-sentence pattern extraction (Krause et al., 2012) and after semantic filtering and enriching have been automatically combined into a single graph. The other resource is *cockrACE*, a specially annotated corpus for the training and evaluation of cross-sentence RE. By employing our powerful annotation tool Recon, annotators mark selected entities and relations (including events), coreference relations among these entities and events, and also terms that are semantically related to the relevant relations and events. This paper describes how the two resources are created and how they complement each other.

Keywords: corpus annotation, coreference resolution, information extraction

1. Introduction

In detecting relation instances within sentences, good results were achieved by distant-supervision RE with seed examples from a knowledge base such as Freebase, finding instances on the web by a search engine and preprocessing of mentions by named-entity detection and dependency parsing. In (Moro et al., 2013), we could demonstrate how the method could be used for n-ary relations and how semantic filtering with lexical-semantics resources such as BabelNet (Navigli and Ponzetto, 2012) effectively boosted precision.

Extending the method to cross-sentence relation mentions poses several challenges. Among them are (i) coreference resolution, (ii) finding instances of semantically related relations that refer to individual arguments and aspects of the target relation and (iii) piecing them together to the appropriate tuples of the target relation. For tackling (i) we are extending an approach by (Xu et al., 2008) that uses domain knowledge and reduces the problem of coreference resolution to the coreferences of immediate relevance for RE. For (ii) we initially work with the rules we acquired from intra-sentential RE since our approach also learns rules or projections of the target relation as long as the main arguments are present. For (iii) we have combined the learned dependency patterns from our rules into a large directed graph termed *sar-graph*, which we will describe in the following section.

The second new type of language resource applies ACE compliant annotation to the marking of kinship relations. All direct and indirect mentions of three next-of-kin relations (marriage, parenthood, sibling) and relevant coreferences for the recognition of these mentions within and across sentence boundaries are annotated in English texts from PEOPLE magazine. The resulting corpus with ACE compliant annotation, which we call *cockrACE* (corpus of coreferences and kinship relations in ACE fashion), will be made freely available to the scientific community for use in research on relation extraction, coreference resolution, textual inference and for any other purpose.

For our own research, cockrACE serves three purposes: Parts of the it will be utilized for establishing a research baseline by measuring recall and precision of RE across sentence boundaries with available means, i.e. with and without sar-graphs and existing coreference resolution in order to evaluate their contributions. A second purpose is the learning of additional patterns needed for cross-sentence RE and for the training of improved specialized coreference resolution. The third application is then not surprisingly the evaluation of the improved RE system by some withheld portion.

2. Sar-graphs

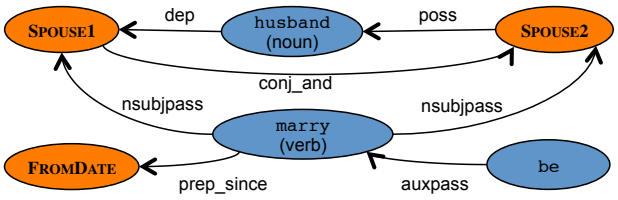
A sar-graph is a graph containing linguistic knowledge at syntactic and lexical semantic levels for a given language and target relation[1]. The nodes in a sar-graph are either semantic arguments of a target relation or content words (to be more exact, their word senses) needed to express/recognize an instance of the relation. The nodes are connected by two kinds of edges: syntactic dependency-structure relations and lexical semantic relations. Thus they are labeled with dependency-structure tags provided by the parser or lexical-semantic relation tags. A formal definition can be found in (Uszkoreit and Xu, 2013).

A sar-graph can be built for every n-ary relation $R\langle a_1, ..., a_n\rangle$ (such as *marriage*: $R\langle$SPOUSE1, SPOUSE2, CEREMONYLOC, FROMDATE, TODATE$\rangle$) and every language l.

Figure 1 shows a simplified sar-graph for two English constructions where *marriage* relations are mentioned. From

[1]The construction of sar-graphs is part of our Google Focused Research Award for Natural Language Understanding: http://googleresearch.blogspot.com/2013/07/natural-language-understanding-focused.html

the two sentences, two dependency patterns can be derived. The relevant trigger words are `husband` and `marry` respectively. These two patterns are connected via their shared semantic arguments, namely, SPOUSE1 and SPOUSE2.



I met Eve's husband Jack.

poss(husband-5, SPOUSE1)
dep(SPOUSE2, husband-5)

Lucy and Peter are married since 2011.

nsubjpass(married-5, SPOUSE2)
conj_and(SPOUSE2, SPOUSE1)
nsubjpass(married-5, SPOUSE1)
auxpass(married-5, are-4)
prep_since(married-5, FROMDATE)

Figure 1: Sar-graph for two English constructions.

The artificially simplified example in Figure 2 may serve to illustrate the structure and contents of a more complex sar-graph. The target relation is *marriage* again. We include only five constructions extracted from the five listed sentences. After each sentence we list the dependency relations from the full parse that belong to the construction. In comparison to the simplified example in Figure 1, this example contains an additional edge type for lexical semantic relations (see the green colored parts). These edges are labeled with semantic relation tags such as hypernym, synonym, troponym, or antonym.

Sar-graphs are constructed automatically by taking the learned dependency-structure patterns for a target relation (Krause et al., 2012) and the lexical semantic network describing a specific relation (Moro et al., 2013) as input. Sar-graphs are a valuable linguistic resource for describing the range of constructions a language offers for expressing an instance of the target relation. So far, we have built sar-graphs for 15 relations[2]. In size they range from 1.4k to 20k vertices and 3.3K to 162k edges each. The sar-graphs also contain many dependency structures that do not always signal instances of the target relation. Instead of filtering these out, we associate them with confidence values determined by our semantic filters and by their positive and negative yield.

3. Annotation Guidelines

There are a number of coreference annotation guidelines proposed in the literature, both for general linguistic phenomena and for the entity and event detection task (Mitkov et al., 2000; Doddington et al., 2004; Linguistic Data Consortium, 2005; Hasler et al., 2006; Komen, 2009). Our approach has built on top of the following two guidelines with some general extension for the cross-sentence relation task:

- ACE annotation guidelines (Doddington et al., 2004; Linguistic Data Consortium, 2005).
- Coreference annotation guidelines proposed by (Komen, 2009)

In comparison to the ACE annotation, we also treat n-ary relations in addition to the binary relations. Furthermore, we also annotate the cross-sentence mentions of arguments contributing to one event or relation mention. Inspired by (Komen, 2009), we classify the coreference relation of noun phrases into the two groups *inferring* and *identity*. In addition, we also annotate lexical-semantic relations among noun phrases through specifications of the *inferring* relation.

The following types of information have been annotated:

- entities: personal entities (name, nominal, pronoun), event entities (name, nominal, pronoun, verb and adjective)
- entity types and subtypes: person, person group, event, date, location
- semantic terms: lexical forms of kinship-relation-relevant word senses (e. g., `marriage`, `sister`)
- relation and event mentions: sentence level and cross-sentence
- coreference relations: source element, target element, identity relation, inferring relation (e.g., synonym, hypernym, hyponym, part-of)
- cross speech relation

We also distinguish *mention extend* and *mention head*. For coreference links, we use the full mention extent of the entity. Given our annotation guidelines, we strive for a fine-grained document level semantic annotation of our target relations.

4. Corpus Setup

In (Krause et al., 2012), we conducted relation-extraction experiments using a corpus of 150 tabloid-press documents, which is a subset of a collection of several thousand PEOPLE-magazine articles from the years 2001–2008. Recently, we published the 150 document corpus along with annotated relation mentions for three semantic relations between people such as *parent-child*, *siblings*, and *marriage* (Li et al., 2014). For the new corpus cockrACE, we selected a different subset from the same base collection of PEOPLE-magazine articles. These 140 documents consist of more than 240,000 words and approximately 8,500 paragraphs, resulting in 1.4 MB. The documents were annotated with mainly three kinds of information: relation mentions for three semantic relations, optionally crossing

[2]These relations are: *acquisition*, *business operation*, *company-product*, *employment tenure*, *foundation*, *headquarters*, *marriage*, *organization alternate name*, *organization leadership*, *organization membership*, *organization relationship*, *organization type*, *parent-child*, *siblings*, and *sponsorship*.

Peter and Lucy exchanged the vows in Paris.

nsubj(exchanged-4, SPOUSE1)
conj_and(SPOUSE1, SPOUSE2)
nsubj(exchanged-4, SPOUSE2)
det(vows-6, the-5)
dobj(exchanged-4, vows-6)
prep_in(exchanged-4, CEREMONYLOC)

I met Eve's husband Jack.

poss(husband-5, SPOUSE1)
dep(SPOUSE2, husband-5)

Lucy and Peter are married since 2011.

nsubjpass(married-5, SPOUSE2)
conj_and(SPOUSE2, SPOUSE1)
nsubjpass(married-5, SPOUSE1)
auxpass(married-5, are-4)
prep_since(married-5, FROMDATE)

I attended the wedding ceremony of Lucy and Peter in 2011.

nn(ceremony-4, wedding-3)
prep_of(ceremony-4, SPOUSE2)
prep_of(ceremony-4, SPOUSE1)
conj_and(SPOUSE2, SPOUSE1)
prep_in(ceremony-4, FROMDATE)

Lucy was divorced from Peter in 2012.

nsubjpass(divorced-3, SPOUSE2)
auxpass(divorced-3, was-2)
prep_from(divorced-3, SPOUSE1)
prep_in(divorced-3, TODATE)

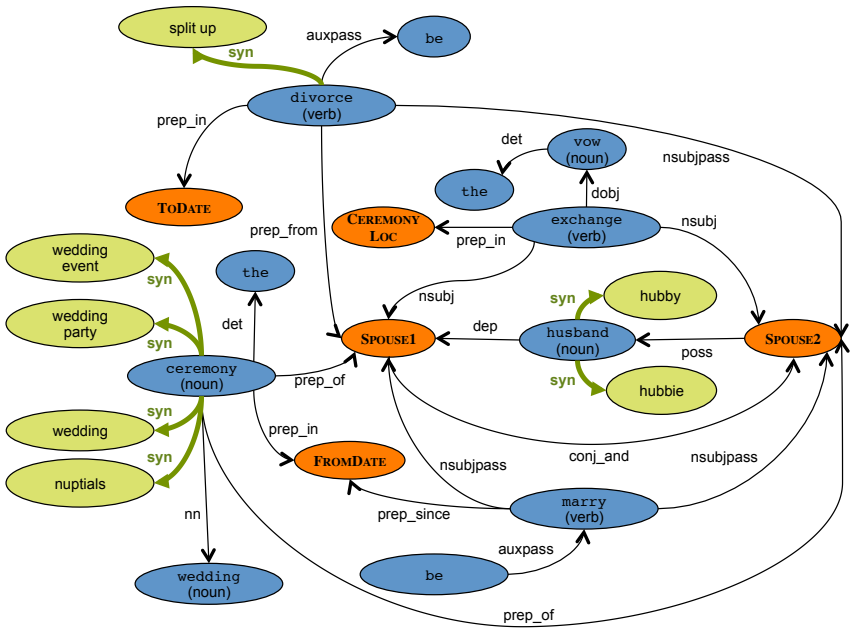


Figure 2: Sar-graph including lexical semantic information.

sentence boundaries, co-referring expressions and lexico-semantically related terms[3].

To speed up the manual annotation, we preprocessed the corpus in the following way:

- **Entity recognition:** Utilizing the NER component of Stanford CoreNLP[4] (Finkel et al., 2005) and a dictionary-based NER module working with Freebase[5] topics, mentions of certain entity types (e. g., person, location) were annotated. In addition, a regular-expression based date recognizer was applied.

- **Sentence segmentation:** We employed the sentence splitter for English from Stanford CoreNLP.

- **Relation mentions:** Using a well performing subset of the extraction patterns from (Moro et al., 2013), we automatically marked potential mentions of the three target relations.

- **Semantic key terms:** In addition to a filter for relation-extraction patterns, Moro et al. (2013) presented automatically learned relation-specific lexical-semantic graphs. We employed the graphs to automatically mark terms which are potential triggers for mentions of the three target relations.

The result of this preprocessing was the annotation of approximately 16,000 sentences, 436 relation mentions, as well as 4,800 mentions of 525 semantic key terms.

[3]We plan to integrate the earlier corpus by carrying out the additional annotations.

[4]http://nlp.stanford.edu/software/corenlp.shtml

[5]http://www.freebase.com

5. Annotation Process

As mentioned above, we follow basically the ACE guidelines and have integrated the proposal of (Komen, 2009) by annotating both identifying and inferring relations between two co-referring expressions. Furthermore, the annotators also annotate semantically related terms that are relevant with respect to the target relation and domain. The latter task implies the consideration of the context and even world knowledge which may give rise to inter-annotator disagreement.

5.1. Annotation Tool: Recon

We have used the annotation tool *Recon* (Li et al., 2012) for marking mentions of concepts and relation mentions in documents. Recon is a Java-based general and flexible annotation tool for annotating n-ary relations among text elements. Compared to other annotation tools, which often only support the annotation of binary relations, Recon allows its user to annotate arbitrary text spans and also n-ary relations between these spans.

Even though the tool had not been utilized before for the annotation of coreferences and lexical semantic relations, it has proven flexible and powerful enough to support these additional types of annotation. Not needing any relation definitions beforehand, the annotator can start right away with marking arbitrary text spans as concept mentions and can assemble these later together with argument-role labels to create semantic-relation mentions. Therefore, this tool is suitable for our free-style annotation tasks, for example building long coreference chains and marking different reference expressions with different semantic labels.

Recon supports the import and export of data in a human-readable and extensive XML format, which facilitates the integration with other NLP tools.

Figures 3 and 4 present Recon screenshots of annotated ex-

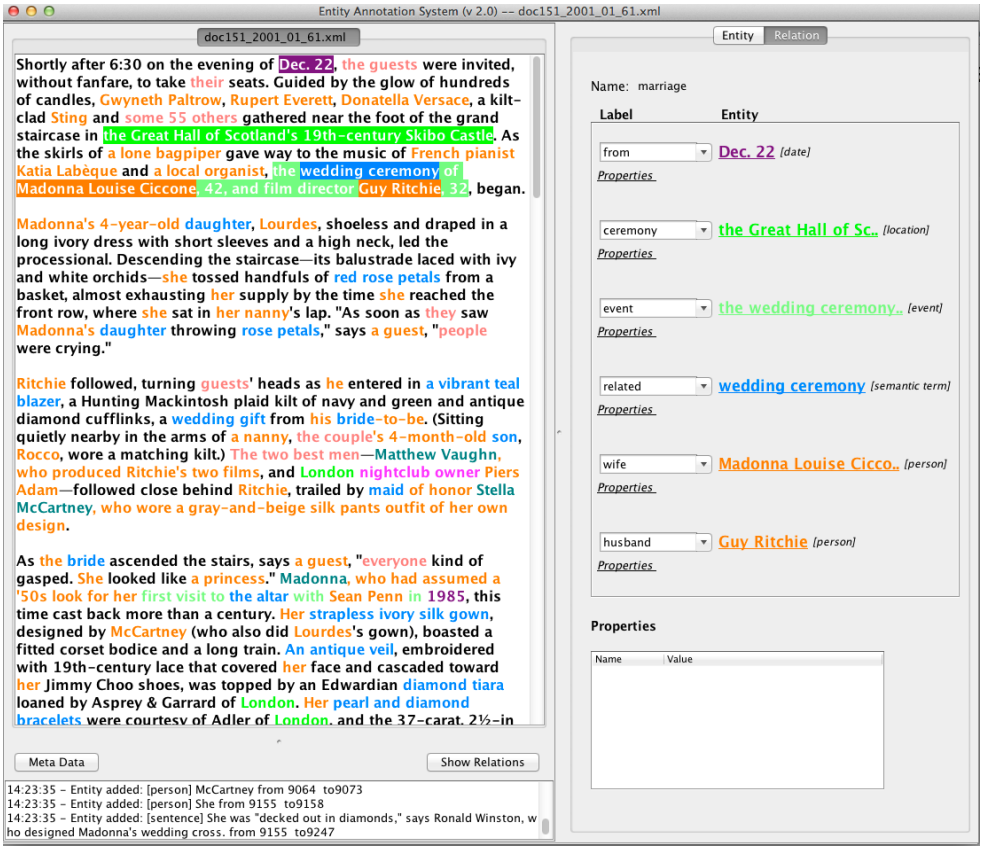


Figure 3: Annotation of cross-sentence relation mention in Recon, cf. Example 1.

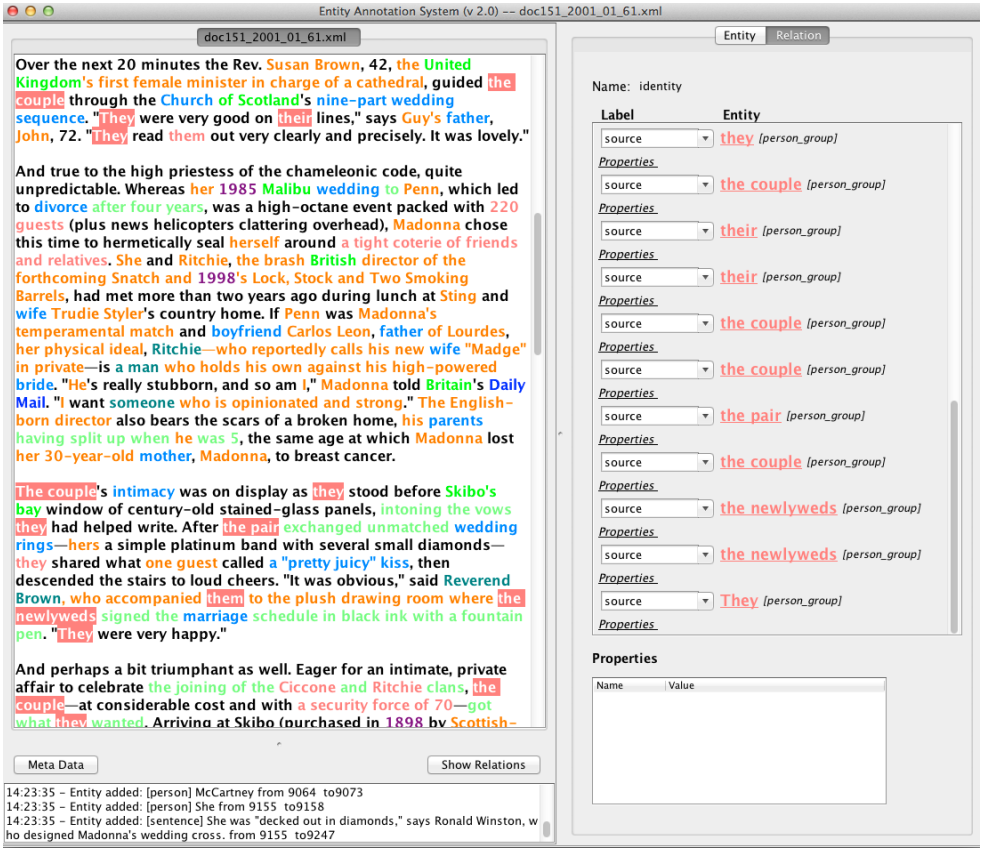


Figure 4: Co-referring noun phrase chain in Recon.

ample documents from the corpus. The first screenshot shows an example for a relation-mention which crosses sentence boundaries. The first paragraph in this example consists of three sentences, all of them mentioning at least one argument of the same marriage relation instance, see Example 1.

Example 1
"Shortly after 6:30 on the evening of Dec. 22, the guests were invited [. . .] to take their seats. [. . .] Sting and some 55 others gathered near the foot of the grand staircase in the Great Hall of Scotland's 19th-century Skibo Castle. [. . .] the wedding ceremony of Madonna Louise Ciccone, 42, and film director Guy Ritchie, 32, began."

Semantic Relation: *marriage*				
Argument	SPOUSE1	SPOUSE2	DATE-FROM	CEREMONY-LOC
Concept Mention	*Madonna Louise Ciccone*	*Guy Ritchie*	*Dec. 22*	*Skibo Castle*

The second figure shows text from the same document as Figure 3, this time with elements of a specific coreference chain being highlighted by a pink box. All of the highlighted expressions refer to the married couple. This example illustrates well why coreferential expressions are such a useful means for cross-sentence relation extraction, because these expressions are spread over the whole text, thus allowing to piece together information bits of the marriage instance from all over the document.
The data is completely annotated from scratch by different annotators. Results of this dual annotation will be compared and discrepancies adjudicated in order to establish inter-annotator agreement scores and identify areas of lingering confusion or inconsistency.

5.2. Statistics and Insights

The overall annotation effort lasted approximately one year. In the beginning, we spent a fair amount of time on discussions with the annotators about relevant literature and published annotation guidelines in order to familiarize them with the different aspects of corpus annotation. We also reserved a subset of the corpus as a basis for discussions on intermediate versions of the annotation guidelines. This slow-paced approach at the start of the annotation effort payed off later on, as there were no major unforeseen obstacles occurring during the annotation process.
After the annotators had finished the annotation of about half of the documents, we set up a discussion to see if changes in the annotation guidelines became necessary. Interestingly, the annotators reported that the results of the automatic preprocessing of the corpus were rather a distraction than a useful support for their task. In particular the annotation of chains of co-referring expressions was too spurious and erroneous to be helpful.
One of the annotators found that certain relation-mention phrases occurred very frequently and that it would be easier to check for the inverse case, i. e., whether instances of such phrases do not consolidate a mention (because of modality in context etc.) rather than marking almost all instances as mentions. In particular, it was suggested to automatically mark certain nested relation-mentions of the form `her mother`, `our sister`, etc. during a post-annotation step in order to speed up the annotation process. We followed this suggestion and modified the guidelines accordingly.
During the post-processing of the annotation we identified some types of annotation errors that could have been prevented if the Recon tool allowed to define task-specific sanity checks. Examples include the specification of allowed entity types per relation argument, allowed argument labels per relation, allowed features per entity/relation type. More advanced functionality such as spelling checks with smarter autocomplete (e. g., `marriiage` should be autocorrected to `marriage`) would improve the annotation as well and facilitate the annotation efforts. We plan to implement such functionality into future versions of the Recon tool.
The annotation process resulted in 45,000 marked concept mentions of the types person, person group, location, and date. Furthermore, more than 1,800 kinship relation mentions have been annotated, about 4,000 sets of co-referring expressions were identified, and approximately 1,300 bridges between singular/plural entity references were resolved. The annotation of one document took the annotators on average 70 to 75 minutes, with a standard deviation of about 20 minutes.

6. Conclusion and Future Work

In this paper, we describe our approach to annotating fine-grained semantic relations among terms, entities, relations and events for document level cross-sentence relation extraction.
We have started to actively use this corpus in our research on relation extraction. Current methods in this area usually focus on individual sentences as their unit of analysis, e.g., extracting only the arguments of a relation instance which are contained within a single sentence (Xu et al., 2007; Moro et al., 2013). Without a corpus like cockrACE progress would be harder to achieve and hardly measurable. As in other areas, we foresee the evolution of both technology and data resources as a bootstrapping process. If cross-sentential RE can be improved with the help of sar-graphs and cockrACE, the enhanced RE system will facilitate the enlargement of the corpus.
Another potential use case for cockrACE is the empirical investigation of coreference and locality phenomena in linguistic research. If cross-sentential RE can be improved with the help of sar-graphs and cockrACE, the enhanced RE system will facilitate the enlargement of the corpus. We are particularly interested in differences between intra-sentential and cross-sentential mentions of n-ary relation instances. Utilizing this corpus, it may be possible to identify a general procedure of how relation-mention arguments are connected by, e. g., chains of co-referring entity mentions or chains of target-relation specific semantic terms.

7. Acknowledgements

This research was partially supported by the German Federal Ministry of Education and Research (BMBF) through the project Deependance (contract 01IW11003) and by

Google through a Focused Research Award granted in July 2013.

8. References

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. The automatic content extraction (ace) program tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*.

Laura Hasler, Constantin Orasan, and Karin Naumann. 2006. NPs for events: Experiments in coreference annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

Erwin R. Komen. 2009. Coreference annotation guidelines. Technical report, Radboud University. `http://erwinkomen.ruhosting.nl/doc/2009_CorefCodingManual_V2-0.pdf`.

Sebastian Krause, Hong Li, Hans Uszkoreit, and Feiyu Xu. 2012. Large-scale learning of relation-extraction rules with distant supervision from the web. In *Proceedings of the 11th International Semantic Web Conference*. Springer, 11.

Hong Li, Xiwen Cheng, Kristina Adson, Tal Kirshboim, and Feiyu Xu. 2012. Annotating opinions in german political news. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Hong Li, Sebastian Krause, Feiyu Xu, Hans Uszkoreit, Robert Hummel, and Veselina Mironova. 2014. Annotating relation mentions in tabloid press. In *Proceedings of the Nineth International Conference on Language Resources and Evaluation (LREC'14)*.

Linguistic Data Consortium. 2005. Annotation guidelines for entity detection and tracking (EDT). Annotation guidelines accompanying the LDC corpus LDC2005T09. `http://catalog.ldc.upenn.edu/docs/LDC2005T09/guidelines/EnglishEDTV4-2-6.PDF`.

Ruslan Mitkov, Richard Evans, Constantin Orasan, Catalina Barbu, Lisa Jones, and Violeta Sotirova. 2000. Coreference and anaphora: developing annotating tools, annotated resources and annotation strategies. In *Proceedings of the Discourse Anaphora and Reference Resolution Conference (DAARC2000)*.

Andrea Moro, Hong Li, Sebastian Krause, Feiyu Xu, Roberto Navigli, and Hans Uszkoreit. 2013. Semantic rule filtering for web-scale relation extraction. In *Proceedings of the 12th International Semantic Web Conference (ISWC 2013)*, Sidney, Australia.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Hans Uszkoreit and Feiyu Xu. 2013. From strings to things – Sar-graphs: A new type of resource for connecting knowledge and language. In *Proceedings of the workshop NLP-DBpedia 2013 (position paper)*, Sidney, Australia.

Feiyu Xu, Hans Uszkoreit, and Hong Li. 2007. A seed-driven bottom-up machine learning framework for extracting relations of various complexity. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 584–591, Prague, Czech Republic, June. Association for Computational Linguistics.

Feiyu Xu, Hans Uszkoreit, and Hong Li. 2008. Task driven coreference resolution for relation extraction. In *Proceedings of ECAI 2008 - 18th European Conference on Artificial Intelligence*, pages 328–332.