

Projection-based Coreference Resolution Using Deep Syntax

Michal Novák and Anna Nedoluzhko and Zdeněk Žabokrtský

Charles University, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské náměstí 25, CZ-11800 Prague 1

{mnovak, nedoluzko, zabokrtsky}@ufal.mff.cuni.cz

Abstract

The paper describes the system for coreference resolution in German and Russian, trained exclusively on coreference relations projected through a parallel corpus from English. The resolver operates on the level of deep syntax and makes use of multiple specialized models. It achieves 32 and 22 points in terms of CoNLL score for Russian and German, respectively. Analysis of the evaluation results show that the resolver for Russian is able to preserve 66% of the English resolver's quality in terms of CoNLL score. The system was submitted to the Closed track of the CORBON 2017 Shared task.

1 Introduction

Projection techniques in parallel corpora are a popular choice to obtain annotation of various linguistic phenomena in a resource-poor language. No tools or gold manual labels are required for this language. Instead, far more easily available parallel corpora are used as a means to transfer the labels to this language from a language, for which such a tool or manual annotation exists.

This paper presents a system submitted to the closed track of the shared task collocated with the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2017).¹ The task was to build coreference resolution systems for German and Russian without coreference-annotated training data in these languages. The only allowed coreference-annotated training data was the English part of the OntoNotes corpus (Pradhan et al., 2013). Alternatively, any publicly available res-

¹Details on the shared task are available in its overview paper (Grishina, 2017) and at <http://corbon.nlp.ipipan.waw.pl/index.php/shared-task/>

olution tool trained on this corpora could be employed.

We adopted and slightly modified an approach previously used by de Souza and Orăsan (2011) and Martins (2015). Parallel English-German and English-Russian corpora are used to project coreference links that had been automatically resolved on the English side of the corpora. The projected links then serve as input data for training a resolver. Unlike the previous works, our coreference resolution system operates on a level of deep syntax. The original surface representation of coreference thus must be transferred to this level. Likewise, coreference relations found by our system must be in the end transformed back to the surface representation, so that they can be evaluated in accordance with the task's requirements. Our resolver also takes advantage of multiple models, each of them targeting a specific mention type.

According to the official results, we were the only participating team. Our system achieved 29.40 points and 30.94 points of CoNLL score for German and Russian portion of the official evaluation dataset, respectively.

The paper is structured as follows. After introducing related works in Section 2, the paper continues with description of the system and its three main stages (Section 3). Section 4 lists the training and testing data to enable evaluation of the proposed system in Section 5. In Section 6, the resolver is analyzed using two different methods. Finally, we conclude in Section 7.

2 Related Work

Approaches of cross-lingual projection have received attention with the advent of parallel corpora. They are usually aimed to bridge the gap of missing resources in the target language. So far, they have been quite successfully applied to

part-of-speech tagging (Täckström et al., 2013), syntactic parsing (Hwa et al., 2005), semantic role labeling (Padó and Lapata, 2009), opinion mining (Almeida et al., 2015), etc. Coreference resolution is no exception in this respect.

Coreference projection is generally approached in two ways. They differ in how they obtain the translation to the language for which a coreference resolver exists. The first approach applies a machine-translation service to create synthetic data in this language. This usually happens at test times on previously unseen texts. Such approach was used by Rahman and Ng (2012) on Spanish and Italian, and by Ogrodniczuk (2013) on Polish.

The other approach, which we employ in this work, takes advantage of the human-translated parallel corpus of the two languages. Unlike the first approach, the translation must be provided already in train time. Postolache et al. (2006) followed this approach using an English-Romanian corpus. They projected manually annotated coreference, which was then postprocessed by linguists to acquire high quality annotation in Romanian. de Souza and Orăsan (2011) applied projection in a parallel English-Portuguese corpus to build a resolver for Portuguese. Our work practically follows this schema, differing in some design details (e.g., using specialized models, resolution on a level of deep syntax). Martins (2015) extended this approach by learning coreference with a specific type of regularization at the end. Their gains over the standard projection come from ability of their method to recover links missing due to projection over inaccurate alignment.

3 System description

Our system for coreference resolution is an example of the projection in parallel corpus. It requires a corpus of parallel sentences in a source (English) and a target language (German and Russian). The procedure consists of three stages illustrated in Figure 1. First, coreference links on the source-language side of the corpus are automatically resolved (see Section 3.1). The acquired links are then projected to the target-language side (Section 3.2). Finally, the target-language side enriched with the projected links is used as a training data to build a coreference resolver (Section 3.3).

3.1 Coreference relations in English

The source-language side of the parallel corpus must get labeled with coreference. In our case, the English side of the parallel corpus already contained annotation of coreference provided by the shared task’s organizers. The annotation is obtained by Berkeley Entity Resolution system (Durrett and Klein, 2014), trained on the English section of OntoNotes 5.0 (Pradhan et al., 2013).

Although Berkeley system is a state-of-the-art performing coreference resolver, we found that it rarely addresses relative and demonstrative pronouns. To label coreference for relative pronouns, we introduced a module from the Treex framework² that employs a simple heuristics based on syntactic trees. Coreference of demonstratives has not been further resolved.

3.2 Cross-lingual projection of coreference

The second stage the proposed schema is to project coreference relations from the source-language to the target-language side of the parallel corpus.

Specifically, we make use of word-level alignment, which allows for potentially more accurate projection. As the parallel data provided for the task are aligned only on the sentence level, word alignment must be acquired on our own. For this purpose, we used GIZA++ (Och and Ney, 2000) a tool particularly popular in the community of statistical machine translation. Even though GIZA++ implements a fully unsupervised approach, which allows for easy extension of the training data with raw parallel texts, it did not prove to be useful for us. We thus obtained word alignment for both the language pairs by running the tool solely on the parallel corpora coming from the organizers.³

Since both German and Russian are morphologically rich languages, we expected word alignment to work better on lemmatized texts. We applied TreeTagger (Schmid, 1995), and MATE tools (Björkelund et al., 2010) for lemmatization in Russian and German, respectively. For robustness, also English texts were preprocessed with a similar procedure, namely a rule-based lemmati-

²Treex (Popel and Žabokrtský, 2010) is a modular NLP framework, primarily designed for machine translation over deep syntax layer. It contains numerous modules for analysis in multiple languages.

³This deserves more experiments with collections of additional data of varying sizes and different algorithms. Due to time reasons we did not manage to finish them, though.

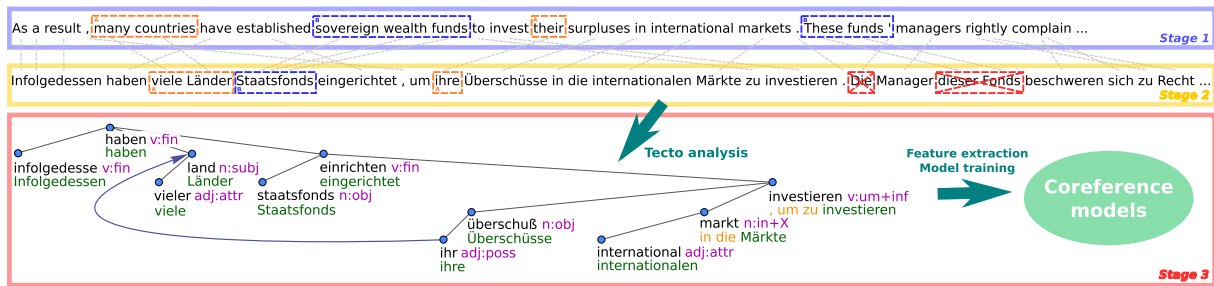


Figure 1: Architecture of the system that consists of three stages: coreference resolution in English (Stage 1), cross-lingual projection of coreference (Stage 2) and resolution in the target language (Stage 3). In the projection stage, the mention *These funds'* is not projected because it is aligned to a discontinuous German span.

zation available as a module in the Treex framework.

Based on the word alignment, the projection itself works as shown in Figure 1. We project mention spans along with its entity identifiers, which are shared among the cluster of coreferential mentions. Only such a mention is projected, whose counterpart forms a consecutive sequence of tokens in the target-language text. In practice, this approach succeeds in projecting around 90% of mentions.⁴

3.3 Coreference resolution in German and Russian

At this point, projected links are ready to serve as training data for a coreference resolver. We make use of an updated version of the already existing resolver implemented within the Treex framework, which operates on a level of deep syntax. All the texts must thus be analyzed and the projected mentions must be transferred up to this level before being used for training.

Analysis up to the tectogrammatical layer.

Treex coreference resolver operates on a level of deep syntax, in Prague theory (Sgall et al., 1986) called tectogrammatical layer. On this layer, a sentence is represented as a dependency tree. Compared to a standard surface dependency tree, the tectogrammatical one is more compact as it consists only of content words (see Figure 1). In addition, several types of ellipsis can be reconstructed in the tree, e.g. pro-drops.

To transform a text in a target language from a surface form to a tectogrammatical representation, we processed it with the following pipelines:

⁴However, they do not need to be necessarily correct, as the alignment may contain errors.

German texts are processed with the MATE tools pipeline (Björkelund et al., 2010) that includes lemmatization, part-of-speech tagging, and transition-based dependency parsing (Bohnet and Nivre, 2012; Seeker and Kuhn, 2012). The surface dependency tree is then converted to the Prague style of annotation using a converter from the HamleDT project (Zeman et al., 2014). Transformation to tectogrammatrics is then performed by a general Treex pipeline, with some language-dependent adjustments.

Russian texts are being parsed directly to the Prague style of surface dependency tree. We trained a UDPipe tool (Straka et al., 2016) on data from SynTagRus corpus (Boguslavsky et al., 2000) converted to the Prague style within the HamleDT project.⁵ Although UDPipe trained on this data is able to lemmatize, we used lemmas produced by TreeTagger instead, as they seemed to be of better quality. In the same fashion as for German, tectogrammatical tree is built from the surface dependency tree using the Treex pipeline adjusted to Russian.

We also included named entity recognition, namely NameTag tool (Straková et al., 2014), to the pipeline. We had trained it on an extended version of the Persons-1000 collection (Mozharova and Loukachevitch, 2016) and named entity annotation of the NoSta-D corpus (Benikova et al., 2014) for Russian and German, respectively.

Transfer of mentions from the surface and back.

On the tectogrammatical layer, a corefer-

⁵We observed lower quality of the Russian parser compared to the German one. The author of UDPipe had instructed us to run the training several times with different values of hyperparameters. However, due to time reasons we ran the training only once, thus probably picking not the most optimal model.

ence link always connects two nodes that represent heads of the mentions. Tectogramatics does not specify a span of the mention, though. The mention usually spans over the whole subtree, except for some notable cases. For instance, an antecedent of a relative pronoun does not include the relative clause itself in its span, even though the clause belongs to a subtree of the antecedent.

The transfer from the surface to the tectogramatics is easy – a head of the mention must be found. We use the dependency structure of a tectogrammatical tree for this and out of all nodes representing nouns or pronouns contained in the mention we pick the one that is closest to the root of the tree.

In the opposite direction, we consider the whole tectogrammatical subtree of a coreferential node. As mentions observed in the datasets rarely include a dependent clause, we rather exclude all such clauses. We skip possible trailing punctuation and finally, we mark the first and the last token of such selection as boundaries of the mention. Due to strict rules to find a mention span and possibly scrambled syntactic parses, this transfer is prone to errors (see Section 6).

Specialized models and features. Treex resolver implements a mention-ranking approach (Denis and Baldridge, 2007). In other words, every candidate mention forms an instance, aggregating all antecedent candidates from a predefined window of a surrounding context. The antecedent candidates are ranked and the one with the highest score is marked as the antecedent. Moreover, a dummy antecedent candidate is added. Highest score for the dummy antecedent implies that the candidate mention is not anaphoric, in fact.

In detail, the resolver consists of multiple models, each of them focused on a specific mention type, e.g., relative pronouns, demonstrative pronouns, or noun phrases. It makes possible to use different windows and different features for each of the types. Personal and possessive pronouns are addressed jointly by two models: a model for personal and possessive pronouns in third person and a model for these pronouns in other persons (in the following denoted as *PP3* and *PPo pronouns*, respectively). Model configurations shared for both languages are listed in Table 1.

Features exploit information collected during the analysis to the tectogrammatical layer. As seen in the table, our models are trained using two kinds

Mention type	DE	RU	Window	Featset
NP	✓	✓	5 prev sents curr sent, preceding	NP
PP3	✓	✓	1 prev sent curr sent, preceding	general
PPo demonstrative	✓	✓		
reflexive	✓	✓	curr sent, all	
reflexive possessive	×	✓		
relative	✓	✓	curr sent, preceding	

Table 1: Configuration of the coreference model for each mention type.

of a feature set:

- *General*: gender and number agreement, other morphological features, distance features, named entity types, syntactic patterns in tectogrammatical trees (to address e.g., relative and reflexive pronouns), dependency relations;
- *NP*: General + head lemma match, head lemma Levehnstein distance, full match; for German: + a similarity score based on word2vec (Mikolov et al., 2013) embeddings⁶ of the mention heads.

The models were trained with logistic regression optimized by stochastic gradient descent. We varied different values of hyperparameters (e.g., number of passes over data, L1/L2 regularization) and picked the setting best performing on the DevAuto set (see Section 4). The learning method is implemented in the Vowpal Wabbit toolkit.⁷

4 Datasets

Raw datasets without manual annotation of coreference are used to train the pipeline described in Section 3. In contrast, manually annotated datasets are reserved exclusively for evaluation purposes. Table 2 shows some basic statistics of the datasets. We refer to each dataset by its label, which consists of two parts. The first part denotes the main purpose of the dataset: *Train* is used for training, *Dev* for development testing, and *Eval* for blind evaluation testing. The second part indicates the origin of the coreference annotation contained in the dataset: *Auto* denotes the projected automatic annotation, *Off* is the official manual annotation provided by the task’s organizers, and *Add*

⁶<http://devmount.github.io/GermanWordEmbeddings/>

⁷https://github.com/JohnLangford/vowpal_wabbit/wiki

Dataset	# Doc.	# Sent.	# EN Tok.	# T Tok.
German				
TrainAuto	4,991	192k	4,834k	4,881k
DevAuto	400	15.4k	391k	395k
DevOff	1	35	–	1k
DevAdd	5	207	5.3k	5.4k
EvalOff	10	404	–	8.8k
Russian				
TrainAuto	3,991	155k	3,847k	3,669k
DevAuto	450	17.5k	436k	417k
DevOff	1	34	–	1k
DevAdd	5	207	5.3k	5.1k
EvalOff	10	412	–	8.1k

Table 2: Statistics of the datasets used throughout this work. The last two columns show the number of tokens in English and in the target language.

denotes the additional dataset annotated by the authors of this paper.

Raw data. We employed the parallel corpora provided by the task’s organizers for building the resolver. Both the English-German and English-Russian corpora come from the *News-Commentary11* collection (Tiedemann, 2012). The datasets were provided in a tokenized sentence-aligned format. We split both corpora into two parts: *TrainAuto* and *DevAuto*. While the former is used for training the models, the latter serves to pick the best values of the learning method’s hyperparameters (see Section 3.3).

Coreference-annotated data. For evaluation purposes, we used two datasets manually annotated with coreference: *DevOff* and *DevAdd*. Except for these datasets, a dataset for the final evaluation (*EvalOff*) of the shared task was provided by the organizer. However, the coreference annotation of this dataset has not been published.

Similarly to the raw data, *DevOff* has been provided by the task’s organizers. In fact, both in German and Russian it is represented by a single monolingual document, presumably coming from the *News-Commentary11* collection.

DevAdd dataset consists of the same five documents randomly selected from both the English-German and English-Russian parallel corpora so that none of these are included in *TrainAuto*. Coreference relations were annotated on all the three language sides. The Russian and English sides were labelled by one of this paper’s co-authors, who speaks native Russian and fluent

Mention type	German		Russian	
	DevOff	DevAdd	DevOff	DevAdd
all	42/370	343/2003	54/497	312/2348
NP	27/312	181/1568	40/475	157/2129
PP3	10/ 16	76/ 142	10/ 10	68/ 70
PPo	0/ 4	33/ 53	1/ 2	29/ 49
demonstrative	1/ 19	9/ 107	0/ 4	0/ 27
reflexive	0/ 7	3/ 48	0/ 1	6/ 9
reflexive possessive	–	–	3/ 5	27/ 29
relative	4/ 12	41/ 85	0/ 0	25/ 35

Table 3: Distribution of mention types in German and Russian coreference-annotated datasets. Denominators show the number of all mention candidates while numerators only of the anaphoric ones.

English, and has long experience of annotating anaphoric relations. The German side was split among three annotators and their outputs were revised by the annotator of the Russian and English part to reach higher consistency. They all followed the annotation guideline published by the organizers.⁸ The reason for creating additional annotated data is that the *DevOff* set consists only of a thousand words per language, which we found insufficient to reliably assess quality of designed systems. The English side was labelled to allow for assessing the quality of the projection pipeline over its stages (see Section 6).

Let us show some notable properties of the German and Russian evaluation data. Table 2 highlights that the *DevAdd* sets expectedly contain five times more words than their *DevOff* counterparts. However, the number of sentences is six times bigger. This may affect a proportion of individual mention types.

Table 3 gives a detailed picture of candidate and anaphoric mentions’ counts. Whereas Russian anaphoric NPs account for 75% of all the anaphoric mentions in *DevOff*, it is only 50% in *DevAdd*. The disproportion appears also between the German datasets.

Finally, some of the mention types appear rarely in the *DevOff* sets. It especially holds for the Russian *DevOff* containing a lack of reflexive, relative and *PPo* pronouns. Conversely, some of the even well-populated types are rarely or never anaphoric (e.g., German demonstrative, reflexive and *PPo* pronouns).

⁸https://github.com/yuliagrishina/CORBON-2017-Shared-Task/blob/master/Parallel_annotation_guidelines.pdf

5 Evaluation

For both German and Russian, we submitted a single system to the shared task. Both the systems fulfill the requirements set on the closed track of the task. To build them we exploited the parallel English-German and English-Russian corpora selected from the News-Commentary11 collection by the task’s organizers.

Metrics. We present the results in terms of four standard coreference measures: MUC (Vilain et al., 1995), B³ (Bagga and Baldwin, 1998), CEAF-e (Luo, 2005) and the CoNLL score (Pradhan et al., 2014). The CoNLL score is an average of F-scores of the previous three measures. It was the main score of some previous coreference-related shared tasks, e.g., CoNLL 2012 (Pradhan et al., 2012), and it remains so for the CORBON 2017 Shared task.

Results. In Table 4, we report the results of evaluating the submitted systems. Comparison across languages shows very similar performance on the DevOff set. However, evaluation on the larger DevAdd set suggests the Russian resolver performs better. Scores on the *EvalOff* dataset confirms higher quality of the Russian resolver, however, the gap is not so big. As the latter dataset is the largest, these results can be considered the most reliable.

6 Discussion

We conducted two additional experiments to learn more about the properties of the projection system. The first experiment investigates the impact of models for individual mention types. The second experiment, in contrast, should tell us more about the quality of the system over its stages.

Model ablations. We conducted a model ablation experiment to shed more light on the model quality and difference between the two evaluation datasets. We repeated the same evaluation, however, each time with a model for a specified mention type left out.

Results in Table 5 show that models for PP3 pronouns and NPs are the most valuable. Better performance of the Russian resolver on DevAdd seems to partly result from a decent model for reflexive possessives, which do not exist in German. Other observations accord with what we highlighted above after inspecting datasets’ statistics

Mention type	German		Russian	
	DevOff	DevAdd	DevOff	DevAdd
all	24.2	22.4	24.2	31.8
⊖ NP	-8.7	-4.6	-7.6	-3.0
⊖ PP3	-11.7	-11.3	-7.5	-10.4
⊖ PPo	+0.5	-1.0	0	-1.1
⊖ demonstrative	+0.5	-0.1	0	0
⊖ reflexive	0	0	0	0
⊖ reflexive possessive	—	—	-4.1	-6.4
⊖ relative	0	-1.9	0	-3.4

Table 5: Results of model ablation. The *all* line describes the complete resolver. Every following line represent an ablated resolver with a model for a given mention type left out. Differences in scores are listed in such line.

in Table 3. There is a big disproportion in score between the two datasets after the model for NPs is removed. This may be a consequence of different ratios of anaphoric NPs to all the anaphoric mentions. Multiple models seem to have marginal, zero, or even negative impact on the final performance. The reasons are threefold:

- low frequency of the mention type in DevOff (e.g., Russian relative and PPo pronouns);
- low frequency of its anaphoric occurrences in the dataset (e.g., all demonstrative pronouns, German reflexive and PPo pronouns)
- the model learned to label most candidates as non-anaphoric (e.g. German demonstrative and reflexive pronouns)

Performance over projection stages. The final performance about 20-30 points seems to be much worse than the CoNLL scores over 60 points observed at the CoNLL 2012 shared task for English. Is coreference resolution in German and Russian so difficult or the projection system deteriorates as it proceeds over its stages?

To answer these questions, we evaluated the output of four stages of the projection CR system. First, we scored the original automatic coreference annotation provided by the Berkeley resolver and the Treex resolver for relative pronouns. This tells us the performance of English CR, which should be comparable with the CoNLL shared task systems. Second, English coreference projected to the target language was evaluated. It should quantify the effect of cross-lingual projection of coreference. Third, all projected coreference relations were transferred to the tectogrammatical layer and

Score	German							Russian						
	DevOff			DevAdd			EvalOff	DevOff			DevAdd			EvalOff
	R	P	F	R	P	F	F	R	P	F	R	P	F	F
MUC	19.0	50.0	27.6	15.7	59.6	24.9	–	16.7	64.3	26.5	23.8	57.5	33.7	–
B ³	13.1	56.1	21.2	11.1	57.6	18.6	–	11.2	71.3	19.3	18.2	56.6	27.5	–
CEAF-e	21.2	27.2	23.8	16.2	44.3	23.7	–	22.1	34.1	26.8	26.9	46.8	34.2	–
CoNLL	24.2			22.4			29.4	24.2			31.8			30.9

Table 4: Evaluation of the resolvers expressed in terms of Precision, Recall and F-score of some popular coreference measures.

back to the surface. This should find the price we pay for conducting coreference resolution at the tectogrammatical layer. Finally, we compare these figures with the final scores presented in Section 5 to see a penalty for modeling coreference.

The experiment was undergone on the DevAdd dataset (see Section 5), annotated with coreference in German, Russian and English. The English part was used to evaluate after the first stage whereas the German and Russian parts for the rest. Performance was measured by CoNLL score.

Figure 2 illustrates how the score declines as the system proceeds over its stages (from left to right). The system for English evaluated after the first stage falls behind the state-of-the-art CR systems by more than 10 points. This can be attributed to a 33-times smaller test set as well as to gentle differences in annotation guidelines. Cross-lingual projection seems to be the bottleneck of the proposed approach. The performance drops by almost 10 points in Russian, even more in German. This could be partially rectified by using better alignment techniques. The loss incurred by operating at the tectogrammatical layer is larger for Russian. It can be attributed to the parsing issues observed on Russian (see Section 3.3). On the other hand, modeling projected coreference by machine learning harms a lot more for German. The models are fit using almost the same feature sets for both languages. Therefore, if the drop is not a consequence of the only difference in features, i.e. word embeddings for German set, it probably results from a different extent of expressive power of the feature set for the two languages. However, this must be taken with a grain of salt as we inferred it without searching for any empirical evidence.

Overall, while our projection-based resolver for Russian is able to preserve 66% of the quality achieved by the English resolver, it is only 46%

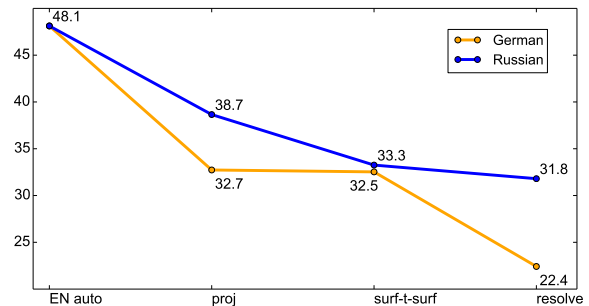


Figure 2: Performance decrease over the projection stages. Measured by CoNLL score.

for German.

7 Conclusion

We introduced a system for coreference resolution via projection for German and Russian. The system does not exploit any manually annotated data in these languages. Instead, it projects the automatic annotation of coreference from English to these languages through a parallel corpus. The resolution system operates on the level of deep syntax and takes advantage of specialized models for individual mention types. It seems to be more suitable for Russian as it is able to achieve 66% of the English resolver’s quality, while it is less than 50% in German, both measured by CoNLL score. We submitted the system to the closed track of the CORBON 2017 Shared task.

Acknowledgments

This project has been funded by the Czech Science Foundation grant GA-16-05394S and the GAUK grant 338915. This work has been also supported and has been using language resources developed and/or stored and/or distributed by the LINDAT/CLARIN project No. LM2015071 of the

Ministry of Education, Youth and Sports of the Czech Republic.

We also gratefully thank to Katja Lapshinova-Koltunski from Saarland University, who helped us with annotation of German texts.

References

- Mariana S. C. Almeida, Cláudia Pinto, Helena Figueira, Pedro Mendes, and André F. T. Martins. 2015. Aligning Opinions: Cross-Lingual Opinion Mining with Dependencies. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, Volume 1: Long Papers*, pages 408–418, Stroudsburg, PA, USA. The Association for Computational Linguistics.
- Amit Bagga and Breck Baldwin. 1998. Algorithms for Scoring Coreference Chains. In *In The First International Conference on Language Resources and Evaluation Workshop on Linguistics Coreference*, pages 563–566.
- Darina Benikova, Chris Biemann, and Marc Reznicek. 2014. NoSta-D Named Entity Annotation for German: Guidelines and Dataset. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Anders Björkelund, Bernd Bohnet, Love Hafdell, and Pierre Nugues. 2010. A High-performance Syntactic and Semantic Dependency Parser. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, pages 33–36, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Igor Boguslavsky, Svetlana Grigorieva, Nikolai Grigoriev, Leonid Kreidlin, and Nadezhda Frid. 2000. Dependency Treebank for Russian: Concept, Tools, Types of Information. In *Proceedings of the 18th Conference on Computational Linguistics-Volume 2*, pages 987–991, Morristown, NJ, USA. Association for Computational Linguistics.
- Bernd Bohnet and Joakim Nivre. 2012. A Transition-based System for Joint Part-of-speech Tagging and Labeled Non-projective Dependency Parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1455–1465, Stroudsburg, PA, USA. Association for Computational Linguistics.
- José G. C. de Souza and Constantin Orăsan. 2011. Can Projected Chains in Parallel Corpora Help Coreference Resolution? In *Proceedings of the 8th International Conference on Anaphora Processing and Applications*, pages 59–69, Berlin, Heidelberg. Springer-Verlag.
- Pascal Denis and Jason Baldridge. 2007. A Ranking Approach to Pronoun Resolution. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1588–1593, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Greg Durrett and Dan Klein. 2014. A Joint Model for Entity Analysis: Coreference, Typing, and Linking. *TACL*, 2:477–490.
- Yulia Grishina. 2017. CORBON 2017 Shared Task: Projection-Based Coreference Resolution. In *The 15th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Vol. 1 (Long Papers)*, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping Parsers via Syntactic Projection Across Parallel Texts. *Natural Language Engineering*, 11(3):311–325.
- Xiaoqiang Luo. 2005. On Coreference Resolution Performance Metrics. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 25–32, Stroudsburg, PA, USA. Association for Computational Linguistics.
- André F. T. Martins. 2015. Transferring Coreference Resolvers with Posterior Regularization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, Volume 1: Long Papers*, pages 1427–1437, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*.
- Valerie Mozharova and Natalia Loukachevitch. 2016. Two-stage Approach in Russian Named Entity Recognition. In *Proceedings of the International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT 2016)*, pages 43–48, Saint Petersburg.
- Franz J. Och and Hermann Ney. 2000. Improved Statistical Alignment Models. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 440–447, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Maciej Ogrodniczuk. 2013. Translation- and Projection-Based Unsupervised Coreference Resolution for Polish. In *Language Processing and Intelligent Information Systems*, number 7912, pages 125–130, Berlin / Heidelberg. Springer.

- Sebastian Padó and Mirella Lapata. 2009. Cross-lingual Annotation Projection of Semantic Roles. *Journal of Artificial Intelligence Research*, 36(1):307–340.
- Martin Popel and Zdeněk Žabokrtský. 2010. TectoMT: Modular NLP Framework. In *Proceedings of the 7th International Conference on Advances in Natural Language Processing*, pages 293–304, Berlin, Heidelberg. Springer-Verlag.
- Oana Postolache, Dan Cristea, and Constantin Orăsan. 2006. Transferring Coreference Chains through Word Alignment. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 889–892, Genoa, Italy. European Language Resources Association.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning - Proceedings of the Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes, EMNLP-CoNLL 2012*, pages 1–40, Jeju, Korea. Association for Computational Linguistics.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. Towards Robust Linguistic Analysis using OntoNotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, Sofia, Bulgaria. Association for Computational Linguistics.
- Sameer Pradhan, Xiaoliang Luo, Marta Recasens, Edward Hovy, Vincent Ng, and Michael Strube. 2014. Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 30–35, Baltimore, Maryland. Association for Computational Linguistics.
- Altaf Rahman and Vincent Ng. 2012. Translation-based Projection for Multilingual Coreference Resolution. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 968–977, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Helmut Schmid. 1995. Improvements In Part-of-Speech Tagging With an Application To German. In *Proceedings of the ACL SIGDAT-Workshop*, pages 47–50, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Wolfgang Seeker and Jonas Kuhn. 2012. Making Ellipses Explicit in Dependency Conversion for a German Treebank. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*, pages 3132–3139, Istanbul, Turkey. European Language Resources Association (ELRA).
- Petr Sgall, Eva Hajičová, Jarmila Panevová, and Jacob Mey. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Springer.
- Milan Straka, Jan Hajič, and Jana Straková. 2016. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4290–4297, Paris, France. European Language Resources Association (ELRA).
- Jana Straková, Milan Straka, and Jan Hajič. 2014. Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 13–18, Baltimore, Maryland. Association for Computational Linguistics.
- Oscar Täckström, Dipanjan Das, Slav Petrov, Ryan T. McDonald, and Joakim Nivre. 2013. Token and Type Constraints for Cross-Lingual Part-of-Speech Tagging. *TACL*, 1:1–12.
- Jörg Tiedemann. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Paris, France. European Language Resources Association.
- Marc Vilain, John Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. 1995. A Model-theoretic Coreference Scoring Scheme. In *Proceedings of the 6th Conference on Message Understanding*, pages 45–52, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. 2014. HamleDT: Harmonized Multi-Language Dependency Treebank. *Language Resources and Evaluation*, 48(4):601–637.