Comparing Rule-Based and Data-Driven Dependency Parsing of Learner Language

Julia Krivanek **Detmar Meurers**
Seminar für Sprachwissenschaft
Universität Tübingen
{krivanek,dm}@sfs.uni-tuebingen.de

Abstract

We explore the performance of two dependency parsing approaches, the rule-based WCDG approach (Foth and Menzel 2006) and the data-driven dependency parser MaltParser (Nivre et al. 2007) on texts written by language learners.

We show that WCDG outperforms MaltParser in identifying the main functor-argument relations, whereas MaltParser is more successful than WCDG in establishing optional, adjunct dependency relations. This can be interpreted as a tradeoff between the rich, hand-crafted lexical resources capturing obligatory argument relations in WCDG and the ability of a data-driven parser to identify optional, adjunct relations based on the linguistic and world knowledge encoded in the gold-standard training corpora.

1 Introduction

Texts written by language learners provide an interesting test case for parsing. They include significant well-formed and ill-formed variation in forms highlighting the robustness of the syntactic analysis performed by different parsing approaches and the resources they use. Dependency parsing is an attractive option in this context, given its focus on the lexical dependency structure serving as interface to interpretation, which avoids further commitments inherent in elaborate constituency-based representations. Parsing learner language is a foundation for any kind of deeper analysis of learner language, as, e.g., needed for automatic content-assessment (Meurers et al. 2011).

Ott and Ziai (2010) describe a dependency parsing experiment based on texts written by American college students learning German. To obtain a gold standard test set, Ott and Ziai (2010) manually annotated this learner corpus using the German dependency annotation scheme developed by Foth (2006) using multiple annotators. For the parsing experiment, they used the data-driven MaltParser (Nivre et al. 2007). They trained this parser on the fifth release of the TüBa-D/Z treebank (Telljohann et al. 2004), after converting it into a dependency treebank format in the style of Foth (2006) using the conversion procedure described in Versley (2005). The TüBa-D/Z treebank consists of newspaper articles, so that there is a significant difference between the training and the test corpus they used. Despite this difference, Ott and Ziai (2010) report that the MaltParser as one of the best current data-driven dependency parsing approaches reliably identified the main functor-argument relation types with a relatively high precision and recall in the 80-90%.

While this is an encouraging result for tasks requiring dependency analysis of learner language, it made us wonder about the impact of the parsing method. In this paper, we therefore explore how parsing of learner language with the data-driven MaltParser compares to parsing with a dependency parser using hand-written rules, for which we make use of the German WCDG parser (Foth and Menzel 2006). Grammar-based parsing with WCDG is based on an information-rich, hand-crafted lexicon (Foth 2006, ch. 2.2). This lead us to hypothesize that the subcategorization requirements hand-coded in the lexicon will contribute to a high-quality coverage of the specific argument requirements of a lexical item. In terms of dependency parsing, this would predict that the rule-based approach in comparison to the data-driven one will fare better in detecting the core functor-argument relations, such as subject and object dependencies. On the other hand, for the subtle distributional differences of adjunct relations, for which relatively few specific constraints are imposed by theoretical linguistic models, a statistical approach trained on corpora – which by their nature encode a combination of language competence, use, and world knowledge – may well fare better.

2 Parsing experiments: The setup

Learner language test corpus We base our parsing experiments on the learner corpus of Ott and Ziai (2010). It is a sub-corpus of the Corpus of Reading Comprehension Exercises in German (CREG, Meurers et al. 2010), which we will refer to as CREG-109. It consists of 109 sentences representing answers to reading comprehension exercises written by US college students at the beginner and intermediate levels of German programs. An example for a learner answer (LA) from the CREG-109 corpus is shown below, where we also show the reading comprehension question (Q) and the teacher's target answer (TA) (but for space reasons not the reading text itself).

Q: *Warum sollte er nicht lachen?*
Why should he not laugh?

TA: *Er sollte nicht lachen, weil das Kind schläft.*
He should not laugh because the baby is sleeping.

LA: *Er sollte nicht lachen für das schlafende Baby.*
He should not laugh for the sleeping baby.

Ott and Ziai (2010) semi-automatically annotated the corpus with STTS part-of-speech tags (Thielen et al. 1999) by running TreeTagger (Schmid 1994) followed by a manual correction phase. On this basis, they manually annotated the corpus according to the dependency annotation scheme devised by Foth (2006), relying on three annotators for each sentence and adjudication for any disagreement to ensure a high quality annotation.

Training MaltParser For data-driven parsing, we essentially followed the setup of Ott and Ziai (2010). We used MaltParser, a system for transition-based dependency parsing (Nivre et al. 2007). The system supports inducing a parsing model from a corpus which has been annotated with dependencies and to parse previously unseen data using the induced model. For **training MaltParser**, we used 90% of the dependency treebank version of the TüBa-D/Z treebank (Telljohann et al. 2004), a corpus consisting of German news texts for which dependency representations in the style of Foth (2006) were obtained with the help of the conversion procedure described in Versley (2005). Training was performed using the LIBSVM learning algorithm and 2-Planar Arc-Eager transition system (Gómez-Rodríguez and Nivre 2010), a linear-time algorithm which is capable of handling limited non-projectivity. The resulting parsing model was used for all MaltParser results reported on in this paper.

Native language test corpus We used sentences from the remaining 10% of the TüBa-D/Z dependency treebank as a **benchmark test corpus** to be able to identify the effect of text type and the impact of parsing learner language in contrast to native language – in line with the well-known fact that parser performance is text type dependent and some text types are more difficult to parse than other ones (Versley 2005). To ensure effective parsing with WCDG, we removed 8% of the sentences, which had character set encoding problems and lexical coverage issues, resulting in a test set of 4142 sentences which we will refer to as TüBa-D/Z test corpus.

The WCDG parser integrates a statistical POS tagger (TnT, Brants 2000) and cannot easily be provided with input including gold standard tags. To make the input to both parsers identical, we thus ran the TnT tagger using the STTS tagset on the CREG-109 and the TüBa-D/Z test corpora and used these automatically tagged version as input for parsing with the MaltParser.

WCDG Parser The WCDG parser representing rule-based dependency parsing in our experiments is an implementation of weighted constraint dependency parsing for German (Foth and Menzel 2006). The WCDG parser allows constraints to express any formalizable property of a dependency tree and the weights for constraints were assigned manually. Parsing with such a WCDG is NP-complete and thus can result in non-termination and efficiency problems. Instead of a full search (*netsearch*) we thus selected *frobbing* as a heuristic search option. Efficiency still remains an issue, so that for our experiments we used the hybrid version of the WCDG parser (Foth and Menzel 2006), which together with a rule-based dependency grammar makes use of a chunker, a supertagger, and a probabilistic shift-reduce parser for labeled dependency trees as stochastic predictor components. While the overall WCDG system successfully tackles parsing of the learner language and the native language test corpus and provides some interesting results, which we now turn to, efficiency clearly is not competitive with statistical dependency parsing, with parse times of several minutes for CREG-109 and several days for the TüBa-DZ test set.

3 Results

3.1 Quantitative evaluation

For the quantitative evaluation we used the *eval.pl* tool from the CoNLL-X shared task on dependency parsing (Buchholz and Marsi 2006). Table 1 sums up the labeled (LAS) and unlabeled (UAS) attachment scores obtained for parsing the CREG-109 and TüBa-DZ test sets with the MaltParser, and Table 2 shows the WCDG parser results.

	LAS	UAS	δ(LAS,UAS)
TüBa-D/Z	84.04%	87.25%	3.21%
CREG-109	78.12%	84.56%	6.44%
			= 3.23% diff.

Table 1: MaltParser results

	LAS	UAS	δ(LAS,UAS)
TüBa-D/Z	81.42%	85.71%	4.29%
CREG-109	79.28%	86.36%	7.08%
			= 2.79% diff.

Table 2: WCDG results

Looking at the results for the learner language test corpus CREG-109, we find that both parsers achieve similar overall results. The WCDG results for parsing the native TüBa-D/Z test corpus for the labeled case are slightly better than for the CREG-109 learner corpus (2.14%), whereas for the unlabeled case the performance for the learner corpus is slightly better (0.65%), probably due to the more complex nature of the TüBa-D/Z news sentences. The linguistic generalizations manually encoded in the WCDG grammar thus appear to be surprisingly applicable to the learner language properties, resulting in a robust parsing performance. MaltParser, on the other hand, with a drop of 5.92% in labeled and 3.31% in unlabeled dependency results between native and learner data shows more clearly that it was trained on the native language news corpus TüBa-D/Z and thereby learned specifics of language and text type which do not generalize that well to reading comprehension answers written by language learners.

An interesting issue arises when one investigates the clear drop between the labeled (LAS) and the unlabeled attachment (UAS) results which arises for both parsers. This drop is significantly larger for the CREG-109 learner corpus than for the native TüBa-DZ corpus, which was also observed by Ott and Ziai (2010) in their CREG-109 parsing experiments with the MaltParser. They hypothesize that this gap may result from the presence of ungrammatical sentences in the corpus.

We investigated this hypothesis for the WCDG parsed CREG-109 corpus by manually inspecting all the relations which were correctly detected but assigned false labels. In other words, we inspected the 53 cases where the parser assigned correct relations but false labels, causing the 7.08% difference between the LAS (79,28%) and UAS (86,36%) results in WCDG parsing CREG-109. We found that 21 of these relations received false labels as the result of an ungrammaticality related to that dependency. We tested this by parsing a corrected version of the sentence with WCDG and observing that the parser then assigned the proper label for the dependency in question. Out of the 53 correctly identified relations with false labels (7.08% of all errors), the 21 cases correspond to 2.8% of all errors which are the result of ungrammaticality. This corresponds exactly to the 2.79% difference in δ(LAS,UAS) between native and learner corpora results of WCDG, fully confirming the hypothesis.

As an example, consider the sentence in (1) and its WCDG parse in Figure 1.

(1) *Sein Eltern hat BA geholfen.*
his parents has BA helped

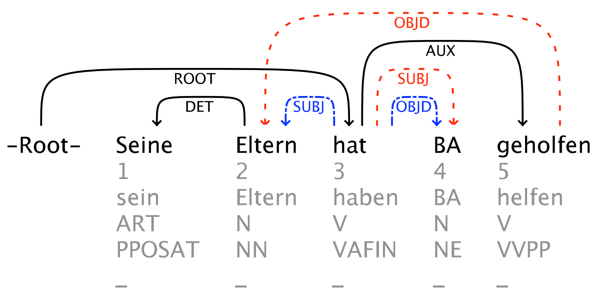

Figure 1: WCDG parser fails to identify subject and object due to subject-verb agreement error

The black solid lines represent correct dependencies identified by the parser, the red dashed dependencies are dependencies incorrectly posited by the parser, and the blue dash-dotted lines are correct dependencies which were not identified by the parser. The learner used a third person singular verb form *hat (has)*, which causes the parser to reject the plural *Eltern (parents)* as subject (despite the singular article *sein (his)*) and to label the dependency with *BA* as the subject instead of as a dative object.

Results for different dependency types To investigate the hypothesis formulated in the introduction that the different parsing approaches will show differences in the way they handle argument compared to the way they handle adjunct relations, we need to take a closer look at the two sets of labeled dependency types. The dependency annotation scheme of Foth (2006) distinguishes a range of argument relations. Given the small size of the CREG-109 corpus, we here focus on the most common ones, for which we have over 10 instances each: SUBJ (subject), OBJA (accusative object), PRED (predicate), and AUX (argument of auxiliary verb). Among the adjuncts, the most common ones are ADV (adverbial modifier) and PP (prepositional adjunct). Note that Foth (2006) uses the labels PP and ADV for grammatical *functions* (adjunct, modifier), different from the typical usage of those labels for grammatical *categories*.

Table 3 shows the results by dependency type for both parsers in percentage figures for precision and recall. The numbers in bold are the best results for a given dependency type.

		MaltParser		WCDG	
Label	#	Recall	Prec.	Recall	Prec.
Argument relations					
SUBJ	95	84.21	80.00	**87.37**	**86.46**
OBJA	52	65.38	70.83	**75.00**	**75.00**
PRED	26	**61.54**	69.57	57.69	**83.33**
AUX	23	60.87	87.50	**73.91**	**94.44**
Modifier relations					
ADV	44	**65.91**	**56.86**	**65.91**	48.33
PP	32	**75.00**	**55.81**	71.88	43.40
Coordination relations					
KON	49	63.27	67.39	**67.35**	**76.74**
CJ	39	82.05	86.49	**89.74**	**92.11**

Table 3: CREG-109 results for the most common argument and adjunct dependency types

We see that in line with our hypothesis, the WCDG parser performs better for each of the lexically subcategorized arguments, the *subject*, *accusative object* and *predicative complements*, and *auxiliary verbal complements* dependencies. The data-driven MaltParser, on the other hand, performs better in identifying *adverbial modifiers* and *prepositional adjuncts*.

The two coordination relations CJ (conjunct) and KON (non-final coordination conjunct) are a special case, because coordinated elements can function as adjuncts or as arguments. We thus manually inspected the coordination relations in the CREG-109 corpus and found that only 3 (about 6 %) of the KON relations in this corpus involve adjuncts. The fact that WCDG parser here outperforms the MaltParser on KON thus also confirms the hypothesis that WCDG is better in detecting argument relations. In the same vein, for the CJ relation the only case where WCDG performed worse than MaltParser is an adjunct case.

The CREG-109 corpus is very small, though, and the small number of instances for each dependency type (shown in the # column) should caution us against overinterpreting these results. On the other hand, we can take a closer look at the parsing results for the larger, native TüBa-D/Z test corpus to see whether we can obtain further support for the interpretation. Table 4 shows the results of parsing the TüBa-D/Z test corpus.

		MaltParser		WCDG	
Label	#	Recall	Prec.	Recall	Prec.
Argument relations					
SUBJ	5408	83.54	87.05	**89.00**	**89.64**
OBJA	2658	75.43	72.96	**79.83**	**82.15**
PRED	1044	**66.48**	71.77	60.82	**76.51**
AUX	2236	85.73	89.41	**91.77**	**96.11**
Modifier relations					
ADV	5115	**78.92**	**77.78**	69.72	64.13
PP	5562	**71.88**	**72.26**	69.67	62.92
Coordination relations					
KON	2531	**76.37**	**71.70**	62.90	71.42
CJ	2164	**90.48**	**91.41**	86.18	83.74

Table 4: TüBa-D/Z results for the most common argument and adjunct dependency types

The table confirms the picture we found for the learner corpus. Again, the WCDG parser obtained the better precision and recall figures for the argument relations (with one exception, the recall of the PRED relation). This is particularly remarkable since we trained the MaltParser on (the 90% development subset of) the TüBa-D/Z cor-

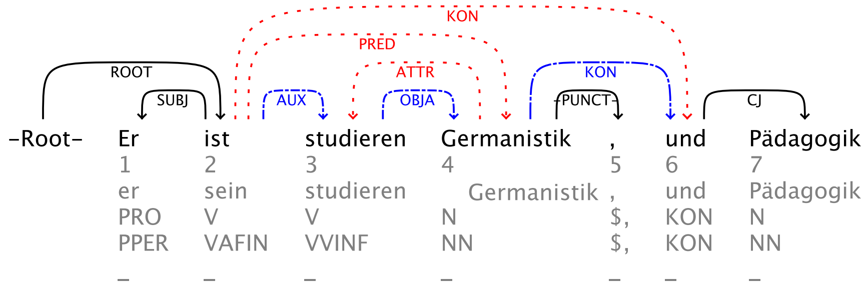

Figure 2: MaltParser: wrong analysis of ill-formed auxiliary verb dependency

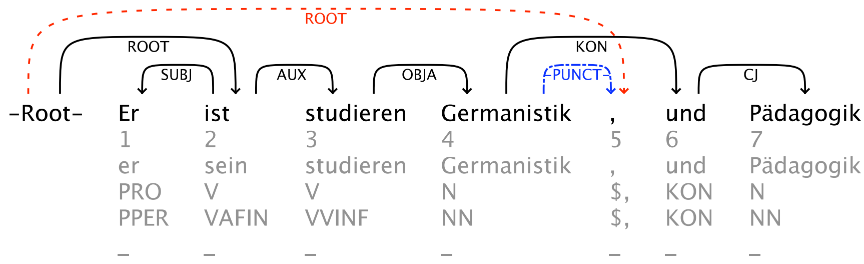

Figure 3: WCDG parser: correct analysis of ill-formed auxiliary verb dependency

pus, which should give it an advantage when parsing the TüBa-D/Z test set. It does indeed improve the results compared to those for the CREG-109, but not enough to overtake the WCDG parser, where the hand-specified lexical subcategorization information apparently is sufficient to maintain an edge. For the modifier relations, on the other hand, the MaltParser significantly outperformed the WCDG parser as expected, confirming the hypothesis that a data-driven parser is better at capturing the characteristics of optional, adjunct relations from those observed in the training data.

The TüBa-D/Z results for the coordination relations KON and CJ, however, are the inverse of the ones we obtained for the CREG-109 corpus. This could be due to the proportion of arguments and adjuncts which are coordinated in the TüBa-D/Z corpus, where the number of coordinated adjuncts is predicted to be higher than in the learner corpus.

3.2 Aspects of a qualitative analysis

Complementing the quantitative analysis, we performed a qualitative inspection of the results obtained for the CREG-109 learner corpus to gain a better understanding of the problems which arise in parsing learner language and how the two parsers differ in this respect.

WCDG: robust parsing of ill-formed AUX An interesting aspect of the results in Tables 3 and 4 is that the scores for identifying arguments of auxiliary verbs are particularly high for the WCDG parser compared to MaltParser, which raises the question why this is the case.

Example (2) illustrates a case where the ungrammatical combination of *ist (is)* with *studieren (study)* prevented the MaltParser from identifying an AUX relation, as shown in Figure 2.

(2) *Er ist studieren Germanistik und Pädagogik.*
he is study German and Pedagogy

The target form of the learner most likely was the English progressive *is studying*, which does not exist as such in German; alternatively, if the learner targeted a perfect tense construction, he chose the wrong auxiliary for this verb and the wrong form for the verbal complement.

Figure 3 shows that the WCDG parser did identify an AUX dependency. Inspection of the WCDG grammar showed that this happens because the WCDG grammar licenses a particular type of passive where the auxiliary *ist* combines with a *zu*-infinitive. For the example (2), the WCDG parser penalized the absence of the particle *zu*, but still this (incorrect) passive analysis achieved the highest weight so that the relation between *ist* and *studieren* was labeled AUX, the most meaningful way to connect these two verbs.

WCDG: robust parsing of subjectless sentences Another interesting issue arises around the analysis of subjects. Example (3) shows an ungrammat-

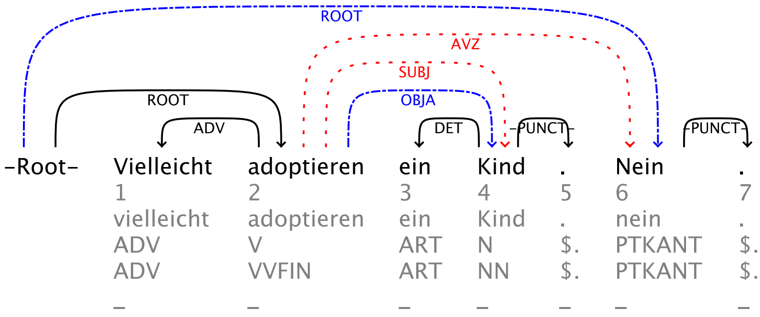

Figure 4: MaltParser: wrong analysis of a sentence missing the subject

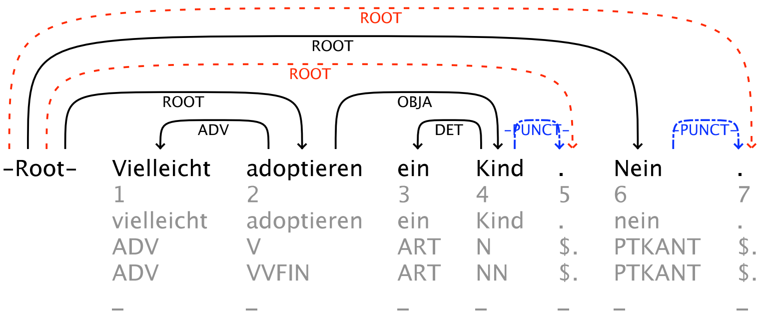

Figure 5: WCDG parser: correct analysis of a sentence missing the subject

ical learner sentence in which the subject is missing (which, different from English, is possible in German in specific cases).

(3) *Vielleicht adoptieren ein Kind. Nein.*
perhaps adopt$_{plur}$ a child no

The analysis of the MaltParser is shown in Figure 4. It posits a subject relation between the finite verb *adoptieren (adopt)* and *Kind (child)*, despite the fact that this relation violates subject-verb agreement between the singular *Kind* and the plural *adoptieren*. Essentially, MaltParser always attempts to identify a subject in a sentence.

Figure 5 shows the WCDG analysis for this sentence. In the WCDG grammar there are restrictions on the cases when a subject can be ordered to the right of a predicate. The sentence under discussion is not among those cases. Thus *Kind* is correctly identified as a object in an overall analysis of a sentence missing a subject.

Naturally, the generalizations captured in the WCDG grammar do not always succeed in balancing the evidence and regularities appropriately for learner language. Example (4), another sentence with a missing subject, is a case in point.

(4) *Rockmusik hören und Mundharmonika spielen*
rock music hear and mouth harp play

Figure 6 shows that in contrast to the sentence we saw in Figure 5, the WCDG parser did not recognize that the subject is missing here and just like the MaltParser labeled the relations between the verbs and the corresponding nouns as subject instead of as objects. The reason for this is a word order rule in the grammar, where a subject in front of a predicate is weighted higher than an object.

4 Conclusion

The aim of this paper was to investigate and compare the performance of a data-driven and a rule-based dependency parsing approach for a learner corpus and a native control corpus. In pursuit of this goal, we reported on parsing experiments with the MaltParser and the WCDG parser for the CREG-109 and a subset of the TüBa-DZ and reported overall results as well as an analysis in terms of the main argument and adjunct relations.

The results highlighted the different strengths of the two parsing approaches. The rich lexical resources of the WCDG parser apparently provide an advantage for identifying the main functor-argument relations, whereas the capability of the supervised machine learning approach of the MaltParser to identify subtle statistical differences in the training data seems to gives it an edge in the analysis of optional, adjunct relations.

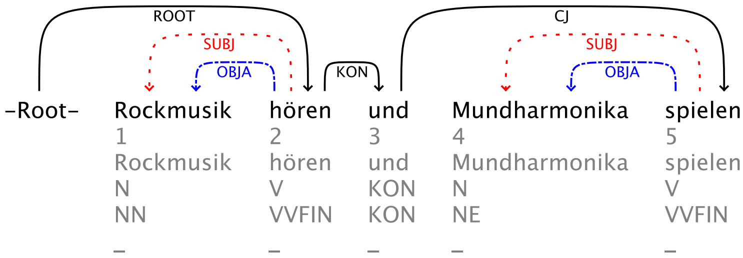

Figure 6: Wrong analysis of both parsers due to word order

In the broader context, this insight essentially lends support to the pursuit of hybrid approaches and parser combinations (e.g., Khmylko et al. 2009; Øvrelid et al. 2009). The learner language domain poses additional challenges to the prioritization of different sources of information is important given that certain language properties are known not to be reliably realized by language learners. While in this paper we have focused on comparing data-driven and rule-based dependency parsing of learner language, an underlying issue which requires more attention is what exactly a dependency analysis of learner language should look like, which has started to receive some attention (Dickinson and Ragheb 2009; Rosén and Smedt 2010; Hirschmann et al. 2010). As far as we see, the criteria crucially depend on the purpose of the analysis, so different types (or multiple layers) of dependency analysis will be needed. On the one hand, a robust dependency analysis glossing over any learner language specifics is needed as a step towards robustly building meaning representations and related processes in applications. On the other hand, detailed dependency analyses based on the various types of evidence that are available when interpreting learner data (morphological, syntactic, and semantic evidence in the data itself, and information about the learner and the task for which the language was produced) could be particularly useful for identifying specific learner language aspects as part of research investigating second language acquisition.

In terms of outlook, while the analysis of the parsing results for the small CREG-109 learner corpus is fully supported by the results obtained for the larger TüBa-DZ test corpus, we would like to extend the analysis to more argument and adjunct relations, for which a larger learner corpus is needed. A larger release of CREG data will become available so that we plan to tackle an extended evaluation based on that larger data set.

Acknowledgments

We would like to thank Yannick Versley, Niels Ott, Ramon Ziai and the anonymous Depling reviewers for their helpful suggestions and discussion.

References

Thorsten Brants, 2000. TnT – A Statistical Part-of-Speech Tagger. In *Proceedings of the Sixth Conference on Applied Natural Language Processing*. Association for Computational Linguistics, pp. 224–231.

Sabine Buchholz and Erwin Marsi, 2006. CoNLL-X shared task on multilingual dependency parsing. In *CoNLL-X '06: Proceedings of the Tenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, Morristown, NJ, USA, pp. 149–164.

Markus Dickinson and Marwa Ragheb, 2009. Dependency Annotation for Learner Corpora. In *Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8)*. Milan, Italy.

Kilian Foth, 2006. *Eine umfassende Constraint-Dependenz-Grammatik des Deutschen*. Technical report, Universität Hamburg.

Kilian A. Foth and Wolfgang Menzel, 2006. Hybrid parsing: using probabilistic models as predictors for a symbolic parser. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACL-44, pp. 321–328.

Carlos Gómez-Rodríguez and Joakim Nivre, 2010. A Transition-Based Parser for 2-Planar Dependency Structures. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. pp. 1492–1501.

Hagen Hirschmann, Anke Lüdeling, Ines Rehbein, Marc Reznicek and Amir Zeldes, 2010. Syntactic Overuse and Underuse: A Study of a Parsed Learner Corpus and its Target Hypothesis. Invited Plenary Talk at the Treebanks and Linguistic Theory Workshop.

Lidia Khmylko, Kilian A. Foth and Wolfgang Menzel, 2009. Co-Parsing with Competitive Models. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*. Association for Computational Linguistics, Paris, France, pp. 99–107.

Detmar Meurers, Niels Ott and Ramon Ziai, 2010. Compiling a Task-Based Corpus for the Analysis of Learner Language in Context. In *Proceedings of Linguistic Evidence*. Tübingen, pp. 214–217.

Detmar Meurers, Ramon Ziai, Niels Ott and Stacey Bailey, 2011. Integrating Parallel Analysis Modules to Evaluate the Meaning of Answers to Reading Comprehension Questions. *IJCEELL. Special Issue on Automatic Free-text Evaluation*, 21(4):355–369.

Joakim Nivre, Jens Nilsson, Johan Hall, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov and Erwin Marsi, 2007. MaltParser: A Language-Independent System for Data-Driven Dependency Parsing. *Natural Language Engineering*, 13(1):1–41.

Niels Ott and Ramon Ziai, 2010. Evaluating Dependency Parsing Performance on German Learner Language. In Markus Dickinson, Kaili Müürisep and Marco Passarotti (eds.), *Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories*. volume 9 of *NEALT Proceeding Series*, pp. 175–186.

Victoria Rosén and Koenraad De Smedt, 2010. Syntactic Annotation of Learner Corpora. In Hilde Johansen, Anne Golden, Jon Erik Hagen and Ann-Kristin Helland (eds.), *Systematisk, variert, men ikke tilfeldig. Antologi om norsk som andrespråk i anledning Kari Tenfjords 60-årsdag [Systematic, varied, but not arbitrary. Anthology about Norwegian as a second language on the occasion of Kari Tenfjord's 60th birthday]*, Novus forlag, Oslo, pp. 120–132.

Helmut Schmid, 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*. Manchester, UK, pp. 44–49.

Heike Telljohann, Erhard Hinrichs and Sandra Kübler, 2004. The TüBa-D/Z Treebank: Annotating German with a Context-Free Backbone. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lissabon.

Christine Thielen, Anne Schiller, Simone Teufel and Christine Stöckert, 1999. *Guidelines für das Tagging deutscher Textkorpora mit STTS*. Technical report, Institut für Maschinelle Sprachverarbeitung Stuttgart and Seminar für Sprachwissenschaft Tübingen.

Yannick Versley, 2005. Parser Evaluation across Text Types. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories (TLT)*. Barcelona, Spain.

Lilja Øvrelid, Jonas Kuhn and Kathrin Spreyer, 2009. Improving data-driven dependency parsing using large-scale LFG grammars. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*. Association for Computational Linguistics, Stroudsburg, PA, USA, ACLShort '09, pp. 37–40.