Why is German Dependency Parsing More Reliable than Constituent Parsing?

Sandra Kübler
Indiana University
skuebler@indiana.edu

Jelena Prokić
Rijksuniversiteit Groningen
j.prokic@rug.nl

1 Introduction

In recent years, research in parsing has extended in several new directions. One of these directions is concerned with parsing languages other than English. Treebanks have become available for many European languages, but also for Arabic, Chinese, or Japanese. However, it was shown that parsing results on these treebanks depend on the types of treebank annotations used [8, 9]. Another direction in parsing research is the development of dependency parsers. Dependency parsing profits from the non-hierarchical nature of dependency relations, thus lexical information can be included in the parsing process in a much more natural way. Especially machine learning based approaches are very successful (cf. e.g. [12, 13]). The results achieved by these dependency parsers are very competitive although comparisons are difficult because of the differences in annotation. For English, the Penn Treebank [11] has been converted to dependencies. For this version, Nivre et al. [14] report an accuracy rate of 86.3%, as compared to an F-score of 92.1 for Charniak's parser [1]. The Penn Chinese Treebank [19] is also available in a constituent and a dependency representations. The best results reported for parsing experiments with this treebank give an F-score of 81.8 for the constituent version [2] and 79.8% accuracy for the dependency version [14]. The general trend in comparisons between constituent and dependency parsers is that the dependency parser performs slightly worse than the constituent parser. The only exception occurs for German, where F-scores for constituent plus grammatical function parses range between 51.4 and 75.3, depending on the treebank, NEGRA [17] or TüBa-D/Z [18]. The dependency parser based on a converted version of Tüba-D/Z, in contrast, reached an accuracy of 83.4% [14], i.e. 12 percent points better than the best constituent analysis including grammatical functions.

Hajič, J. and Nivre, J. (eds.): Proceedings of the TLT 2006, pp. 7–18.

In this paper, we will examine reasons for this difference in quality in parsing German that was found previously. Our hypothesis is that the dependency parser is capable of analyzing especially long-distance relationships and coordination phenomena better. In order to validate our hypothesis, we selected samples of sentences displaying the above mentioned phenomena, parsed them with a constituent parser and a dependency parser, and conducted an error analysis for the two versions. For this experiment, we used the TüBa-D/Z treebank, LoPar [16] as constituent parser, and MaltParser [14] as dependency parser. In the following sections, we will first provide a short overview of the annotation scheme used for TüBa-D/Z and the dependency version of the treebank. Then we will describe the experimental setup and the results from these experiments and draw our conclusions.

2 The Two Versions of the Treebank TüBa-D/Z

2.1 The Constituent Treebank

TüBa-D/Z in its original form is a constituency-based treebank enriched with function-argument structure. The treebank is based on the German newspaper, 'die tageszeitung'. The version that was used for the experiments comprised 22 000 sentences[1]. The TüBa-D/Z annotation scheme distinguishes four levels of syntactic constituency: the lexical level, the phrasal level, the level of topological fields, and the clausal level. The primary ordering principle of a clause is the inventory of topological fields, which characterize the word order regularities among different clause types of German, and which are widely accepted among descriptive linguists of German (cf. e.g. [3, 5]). The TüBa-D/Z annotation relies on a context-free backbone (i.e. proper trees without crossing branches) of phrase structure combined with edge labels that specify the grammatical functions of constituents. The annotation scheme of the TüBa-D/Z is described in more detail in [18].

(1) Der Autokonvoi mit den Probenbesuchern fährt eine Straße entlang, die noch
The car convoy with the rehearsal visitors goes a street along, which even
heute Lagerstraße heißt
today Lagerstraße is called.
'The convoy of the rehearsal visitors' cars goes along a street that is still called Lagerstraße.'

Figure 1 shows a typical example from TüBa-D/Z, for sentence (1). The POS annotation is based on the STTS tagset [15]; it is shown under the words. Syntactic categories are annotated as nodes, grammatical functions as edges shown as

[1] In the meantime, a new release was made available with 27 000 sentences.

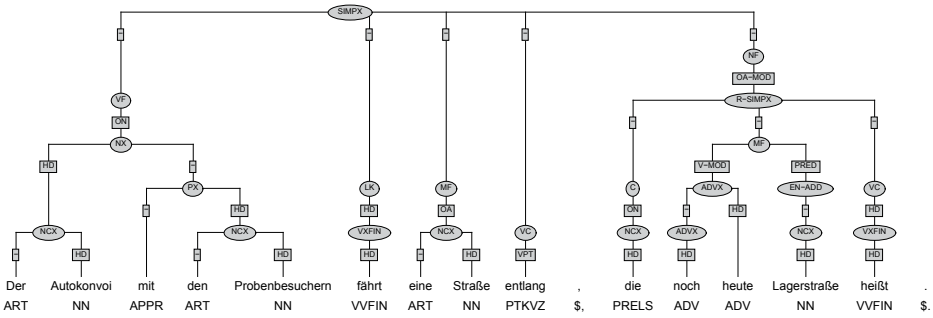

Figure 1: A sample tree from the TüBa-D/Z treebank.

square boxes. The noun phrase *Der Autokonvoi mit den Probenbesuchern* is the subject (ON), the noun phrase *eine Straße* the accusative object (OA) of the main clause. The relative clause consists of the relative pronoun as the subject, an adverbial phrase that modifies the verb (V-MOD) and the predicate noun phrase (PRED). The relative clause modifies the accusative object of the main clause, this is expressed by the label OA-MOD rather than by a crossing branch. In addition to the constituents and the grammatical functions, the sentence is grouped into topological fields: The main verb constitutes the head (HD) of the sentence, the subject is placed in the initial field (VF), the accusative object constitutes the middle field (MF), the verbal particle is placed in the verb complex (VC), and the relative clause in the final field (NF).

2.2 The Dependency Version

The constituent treebank described above has been converted into dependencies. The target dependency annotation is based on the annotation used for the Constraint Dependency Parser [4]. The transformation is based on the grammatical function information and heuristics in such cases where the constituent annotation does not provide enough evidence. The two phenomena that are of most interest for our study are long-distance relationships and coordination. Long-distance relationships, marked by specific functional labels in the constituent tree, are converted into non-projective dependencies. The dependency annotation for the sentence in Figure 1 is shown in Figure 2. Note that the extraposed relative clause is now directly dependent on the noun *Straße*.

In coordinations, the first conjunct is treated as the head of the coordination, the conjunction is dependent on the first conjunct, and the second conjunct on the coordination. Sentence (2) gives an example of a coordinated noun phrase. The constituent annotation is shown in Figure 3, the dependency representation in

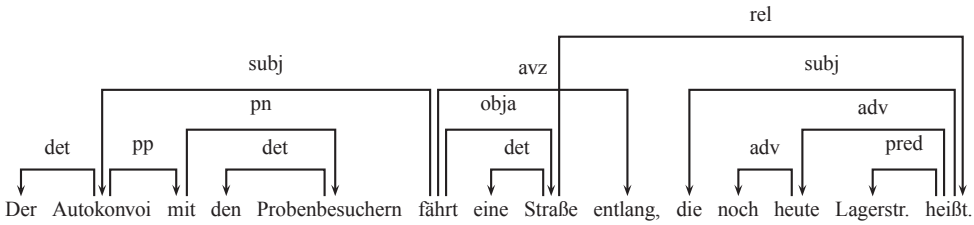

Figure 2: The dependency representation of the tree in Figure 1. The street name is shortened for representational purposes.

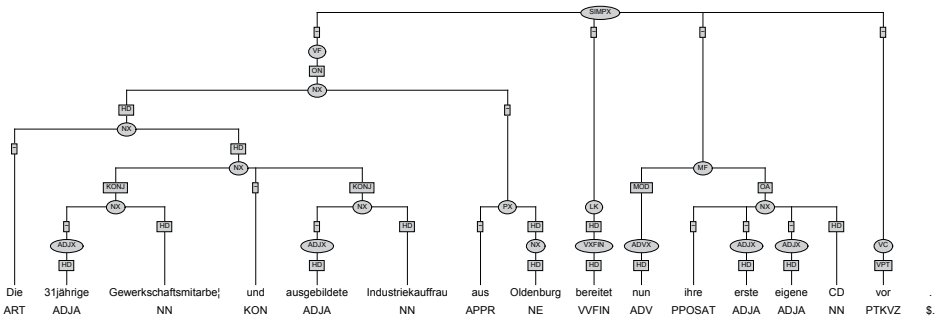

Figure 3: A sentence containing a coordinated noun phrase.

Figure 4.

(2) Die 31jährige Gewerkschaftsmitarbeiterin und ausgebildete Industriekauffrau
The 31 year old union member and trained industrial clerk
aus Oldenburg bereitet nun ihre erste eigene CD vor.
from Oldenburg prepares now her first own CD *part.*
'The 31 year old union member and trained industrial clerk from Oldenburg now prepares her first CD.'

3 The Parsing Experiment

Our hypothesis is that the difference in parsing results is, among other reasons, due to the better handling of coordination and long-distance phenomena in dependency parsing. For this reason, we concentrated the evaluation on samples from the treebank which contain such phenomena. We considered the following phenomena: For long-distance phenomena, we concentrated on different types of fronted modifiers and on a comparison of extraposed relative clauses and adjacent relative clauses. For coordination, different types of coordination were chosen: phrasal coordination, clause coordination, topological field coordination, and asymmetrical

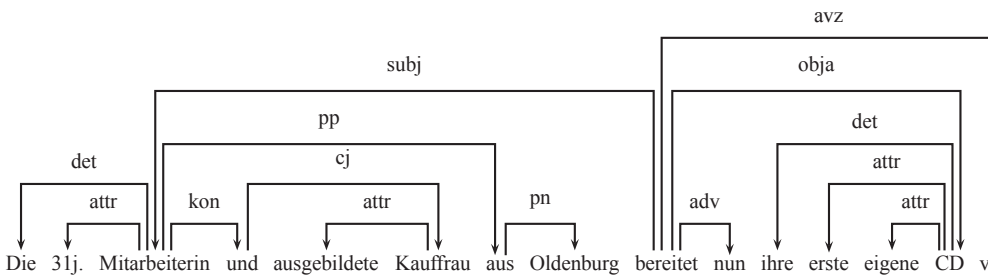

Figure 4: The dependency representation of the (shortened) sentence (2).

coordination.

For the experiments, a 10-fold cross-validation setup was used, and the parsers had access to the gold POS tags for the words. From the parsers' outputs, we then extracted a predefined set of sentences for evaluation from both the constituent parses and the dependency parses. Thus, for each sentence, both analyses were compared. For evaluation, these parsed sentences were grouped into the following groups: 1) correctly parsed, 2) wrong attachment site, 3) wrong label, 4) wrong boundaries, and 5) no parse. The first group contains sentences that were parsed correctly concerning the phenomenon in question. Thus, a sentence containing an extraposed relative clause was considered correct if the relative clause was grouped with the correct boundaries and the correct label, and if it was attached to the correct constituent or word. Whether the parse contained mistakes concerning the internal structure of the relative clause has no effect on the evaluation. The second group contains sentences for which the parser found the correct boundaries and labels but attached the constituent to the wrong constituent in the tree or to the wrong word in the dependency graph. The third group contains sentences where the constituent or dependency under inspection received the wrong label. If the constituents were recognized with an incorrect yield, they are assigned to the fourth group. And if the constituent or dependency is missing completely or if the sentences did not receive any parse, the sentence is assigned to the fifth group.

4 Results

This section presents results from the evaluation of the phenomena listed in the previous section. The figures in this evaluation, however, must be interpreted in the light of the selection process. We have consciously selected sentences containing phenomena that are notoriously difficult for parsers to analyze. Additionally, these sentences are extracted from newspaper text so that they often exhibit a complex

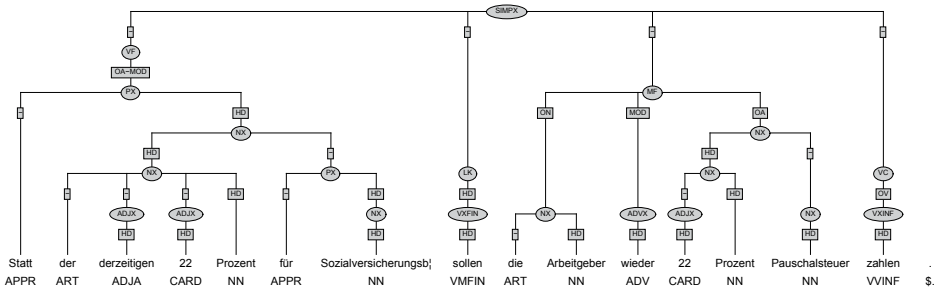

Figure 5: A long-distance relationship involving a fronted modifier (OA-MOD) of the accusative object (OA).

structure apart from the phenomena that are under investigation in this study. Thus, the following evaluation does in no way give any insight into the overall quality of the parsers.

4.1 Fronted Modifiers

In German main clauses, there is the restriction that exactly one constituent occurs in front of the finite verb, i.e. in the initial field. This constituent is the subject in approximately 50% of the cases. There are, however, also cases, in which the initial field is occupied by a modifier of one of the constituents in the middle field. An example for such a sentence is given in (3), the constituent analysis is shown in Figure 5.

(3) Statt der derzeitigen 22 Prozent für Sozialversicherungsabgaben sollen die
Instead of the present 22 percent for social security contributions should the
Arbeitgeber wieder 22 Prozent Pauschalsteuer zahlen.
employers again 22 percent lump-sum tax pay.
'Instead of paying 22 percent for social benefits, the employers are to pay 22 percent as a general tax.'

For this evaluation, we selected sentences whose initial fields are occupied by modifiers that modify different constituents in the middle field. We selected 10 sentences each that modify the subject (ON) and the accusative object (OA). For modifiers of other constituents, all occurrences in the treebank were used: modifiers of the dative object (OD; 2 sentences), of the predicate (PRED; 6 sentences), and of another modifier (MOD; 6 sentences). The results of this evaluation are shown in Table 1. For parsers, the major problem in the analysis of such fronted constituents is the decision which constituents they modify. Both parsers made the majority of their errors in this category. It is, however, clear that the dependency parser parsed

modifier	rep.	corr.	wr. attach.	wr. label	wr. bound.	no parse
ON-MOD	const.	0	9	0	0	1
	dep.	3	7	0	0	0
OA-MOD	const.	0	9	1	0	0
	dep.	0	9	0	0	1
OD-MOD	const.	0	2	0	0	0
	dep.	0	2	0	0	0
PRED-MOD	const.	0	6	0	0	0
	dep.	4	2	0	0	0
MOD-MOD	const.	1	4	0	0	1
	dep.	4	2	0	0	0

Table 1: The evaluation of fronted modifiers.

nearly one third of these fronted constituents correctly while the constituent parser succeeded in only one case.

4.2 Relative Clauses

In German, relative clauses can be attached directly to the noun phrase that they modify, or they can be extraposed, resulting in a long-distance relationship. As described above, this is annotated via a special functional label in the constituent representation. In the dependency annotation, this relationship is annotated via a non-projective dependency. An example for such a relative clause is given in Figures 1 and 2. For the present evaluation, we selected 20 sentences for each type of relative clauses. The results of this evaluation are shown in Table 2. These results show, as expected, that relative clauses that are attached directly to the noun phrase can be recognized more reliably. The constituent parser correctly groups 6 of these relative clauses, the dependency parser 12. However, if the parsers do not analyze the relative clauses correctly, the errors that they make are rather severe. For 6 more sentences, the constituent parser either did not produce any parse or a parse which contained no relative clause. For sentence (4), the verb of the relative clause was analyzed as the verb complex of the main clause, and the three phrases *an dem*, *Max Daniel*, and *Professor* were grouped into a single prepositional phrase. The dependency parser analyzed the relative clause correctly but attached it to the noun *Gesang*.

	adjacent rel. cl.		extraposed rel. cl.	
error type	const.	dep.	const.	dep.
correct	6	12	0	9
wrong attachment	1	4	0	4
wrong label	3	0	18	7
wrong boundaries	4	1	1	0
no parse	6	3	1	0

Table 2: The evaluation of relative clauses.

(4) Renato Mismetti hat zunächst Psychologie studiert und dann Gesang an der
Renato Mismetti has at first psychology studied and then voice at the
Universität Uberlandia in Brasilien, an der Max Daniel Professor für Klavier ist.
University Uberlandia in Brasil, at which Max Daniel professor for piano is.
'Renato Mismetti first studied psychology then voice at the University of Uberlandia in Brasil where Max Daniel is piano professor.'

For the extraposed relative clauses, the constituent parser recognized 18 out of the 20 examples as subordinate clauses rather than as relative clauses. The remaining two sentences were analyzed in a similar way to example (4), i.e. no clausal constituent was found. The dependency parser correctly analyzed 9 sentences; for another 7 sentences, the relative clauses were analyzed correctly, but attached to the wrong noun or to a verb.

4.3 Coordination

From the many different phenomena involving coordination, we chose a selection, which covers a range of complexity. We start out with a group of sentences which contain coordination on the phrasal level, including 10 sentences with noun phrase coordination (`NX`) and 10 sentences with adjectival phrase coordination (`ADJX`). Then there is a group of 10 sentences involving coordination on the clause level. The next group involves the coordination of combinations of topological fields. In this case, the node dominating the conjuncts is labeled `FKOORD`, the conjuncts are annotated with the edge label `FKONJ` in constituent annotation. This group also contains 10 sententes with a subject gap in the second conjunct. An example for this phenomenon can be found in (5). Here, the subject *einer* is only present in the first conjunct *kommt einer* but not in the second *stiehlt mir meine Krise*. In the dependency representation, this results in the subject only being dependent on the first verb.

coord. type	rep.	corr.	wr. attach.	wr. label	wr. bound.	no parse
phrase c.	const.	10	0	3	7	0
	dep.	12	1	3	4	0
clause c.	const.	8	0	0	2	0
	dep.	8	0	0	2	0
field c.	const.	11	0	0	8	1
	dep.	16	0	1	2	1
asymm. c.	const.	1	0	1	14	4
	dep.	3	0	0	14	3

Table 3: The evaluation of coordination phenomena.

(5) Immer kommt einer und stiehlt mir meine Krise.
Always comes someone and steals me my crisis.
'Every time someone comes and steals my crisis.'

The next category contains cases of asymmetrical coordination. In such cases, the syntactic category dominating the conjuncts in the constituent representation is the syntactic category of the first conjunct. In (6), the asymmetrical coordination concerns the noun phrase *nur noch Außenminister* and the adjectival phrase *nicht mehr grün*. In this case, the constituent that dominates the coordination is annotated as a noun phrase.

(6) Falls die Delegierten ihm die Gefolgschaft verweigern, wird befürchtet, daß der
If the delegates him the allegiance deny, is feared that the
grüne Außenminister nur noch Außenminister und nicht mehr grün ist.
green foreign minister only still foreign minister and no more green is.
'If the delegates deny him their allegiance, there is the fear that the green foreign minister will still be foreign minister but not green anymore.'

The results of this evaluation are shown in Table 3. This table shows that coordination is equally difficult for both parsers. While phrasal and clausal coordination can be handled by the parsers fairly reliably, especially asymmetrical cases are very error-prone. Most of the errors here are due to the attempts of the parsers to analyze the coordination as symmetrical. The fact that the dependency parser parsed 16 out of the 20 sentences with field coordination correctly may be a result of the fact that field information is not explicitly annotated in the dependency trees. However, it would be wrong to argue that consequently field information should be deleted from the constituent trees. Previous studies [7, 10] have shown that this deletion has a detrimental effect on constituent parsing.

5 Conclusion and Future Work

In this study, we investigated the qualitative differences in constituent and dependency parsing for the German treebank Tüba-D/Z. Our hypothesis is that the dependency parser performs better because it can handle long-distance relationships and coordination better. The in-depth analysis of parsed sentences which exhibit these phenomena shows that our assumption is valid. For fronted modifiers, for extraposed relative clauses, and for different coordination phenomena, the number of correct analyses of the dependency parser is consistently higher than for the constituent parser. One explanation for these differences in performance can be sought in the architecture of the specific dependency parser, MaltParser. This parser employs a definition of a variable context, which is accessible when decisions are made. When a word is analyzed as a dependent of a previous word, it is removed from the immediate context, so that the next word moves into the context. However, this can only be a partial explanation since the dependency parser also performs better in the cases of adjacent relative clauses and of phrase coordination, two phenomena, which require less context to be analyzed. This leads to the hypothesis that another important factor is the use of lexical information in the dependency parser.

For the future, we are planning to repeat the dependency experiments without giving the parser access to lexical information. If the results are comparable to the results presented here, we can conclude that the flexible context is more important than lexical information. Additionally, we would like to extend the experiments to include also the other German treebank, TIGER, for which there also exists a dependency version. And we are planning to use different parsers, the MSTParser [12] and the Stanford parser [6], in order to investigate if the results are due to the selection of the parsers used in this experiment or whether our hypothesis is independent of the specific parsing algorithm.

Acknowledgments

We are grateful to Yannick Versley, who converted TüBa-D/Z to dependencies, and to Wolfgang Maier, Joakim Nivre, and Tylman Ule for their evaluation programs. This research was supported by the German Research Council in the framework of the SFB 441, Project A1.

References

[1] Charniak, Eugene (2000) A maximum-entropy-inspired parser. In *Proceed-*

ings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics and the 6th Conference on Applied Natural Language Processing, ANLP/NAACL'2000, pp. 132–139, Seattle, WA.

[2] Chiang, David and Bikel, Daniel (2002) Recovering latent information in treebanks. In *Proceedings of the 19th International Conference on Computational Linguistics, COLING-2002*, Taipei, Taiwan.

[3] Drach, Erich (1937) *Grundgedanken der Deutschen Satzlehre.* Frankfurt/M.: Diesterweg.

[4] Foth, Kilian, Daum, Michael and Menzel, Wolfgang (2004) A broad-coverage parser for German based on defeasible constraints. In *KONVENS 2004, Beiträge zur 7. Konferenz zur Verarbeitung natürlicher Sprache*, pp. 45–52, Vienna, Austria.

[5] Höhle, Tilman (1986) Der Begriff "Mittelfeld", Anmerkungen über die Theorie der topologischen Felder. In *Akten des Siebten Internationalen Germanistenkongresses 1985*, pp. 329–340, Göttingen, Germany.

[6] Klein, Dan and Manning, Christopher (2003) Accurate unlexicalized parsing. In *Proceedings of ACL-2003*, pp. 423–430, Sapporo, Japan.

[7] Kübler, Sandra (2005) How do treebank annotation schemes influence parsing results? Or how not to compare apples and oranges. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2005*, pp. 293–300, Borovets, Bulgaria.

[8] Kübler, Sandra, Hinrichs, Erhard W. and Maier, Wolfgang (2006) Is it really that difficult to parse German? In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP 2006*, pp. 111–119, Sydney, Australia.

[9] Levy, Roger and Manning, Christopher (2003) Is it harder to parse Chinese, or the Chinese treebank? In *Proceedings of ACL 2003*, pp. 439–446, Sapporo, Japan.

[10] Maier, Wolfgang (2006) Annotation schemes and their influence on parsing results. In *Proceedings of the ACL-2006 Student Research Workshop*, Sydney, Australia.

[11] Marcus, Mitchell, Santorini, Beatrice and Marcinkiewicz, Mary Ann (1993) Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313–330.

[12] McDonald, Ryan, Crammer, Koby and Pereira, Fernando (2005) Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pp. 91–98, Ann Arbor, MI.

[13] Nivre, Joakim, Hall, Johan and Nilsson, Jens (2004) Memory-based dependency parsing. In *Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL)*, pp. 49–56, Boston, MA.

[14] Nivre, Joakim, Hall, Johan, Nilsson, Jens, Chanev, Atanas, Eryiğit, Gülşen, Kübler, Sandra, Marinov, Svetoslav and Marsi, Erwin (To appear) MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*.

[15] Schiller, Anne, Teufel, Simone and Thielen, Christine (1995) Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical report, Universität Stuttgart and Universität Tübingen.

[16] Schmid, Helmut (2000) LoPar: Design and implementation. Technical report, Universität Stuttgart.

[17] Skut, Wojciech, Krenn, Brigitte, Brants, Thorsten and Uszkoreit, Hans (1997) An annotation scheme for free word order languages. In *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP)*, Washington, D.C.

[18] Telljohann, Heike, Hinrichs, Erhard W., Kübler, Sandra and Zinsmeister, Heike (2005) *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z)*. Seminar für Sprachwissenschaft, Universität Tübingen, Tübingen, Germany.

[19] Xue, Naiwen, Xia, Fei, Chiou, Fu-Dong and Palmer, Martha (2005) The Penn Chinese TreeBank: Phrase structure annotation of a large corpus. *Natural Language Engineering*, 11(2), 207–238.