The Benefit of Stochastic PP Attachment to a Rule-Based Parser

Kilian A. Foth and Wolfgang Menzel
Department of Informatics
Hamburg University
D-22527 Hamburg
Germany
foth|menzel@nats.informatik.uni-hamburg.de

Abstract

To study PP attachment disambiguation as a benchmark for empirical methods in natural language processing it has often been reduced to a binary decision problem (between verb or noun attachment) in a particular syntactic configuration. A parser, however, must solve the more general task of deciding between more than two alternatives in many different contexts. We combine the attachment predictions made by a simple model of lexical attraction with a full-fledged parser of German to determine the actual benefit of the subtask to parsing. We show that the combination of data-driven and rule-based components can reduce the number of all parsing errors by 14% and raise the attachment accuracy for dependency parsing of German to an unprecedented 92%.

1 Introduction

Most NLP applications are either data-driven (classification tasks are solved by comparing possible solutions to previous problems and their solutions) or rule-based (general rules are formulated which must be applicable to all cases that might be encountered). Both methods face obvious problems: The data-driven approach is at the mercy of its training set and cannot easily avoid mistakes that result from biased or scarce data. On the other hand, the rule-based approach depends entirely on the ability of a computational linguist to anticipate every construction that might ever occur. These handicaps are part of the reason why, despite great advances, many tasks in computational linguistics still cannot be performed nearly as well by computers as by human informants.

Applied to the subtask of syntax analysis, the dichotomy manifests itself in the existence of *learnt* and *handwritten* grammars of natural languages. A great many formalisms have been advanced that fall into either of the two variants, but even the best of them cannot be said to interpret arbitrary input consistently in the same way that a human reader would. Because the handicaps of different methods are to some degree complementary, it seems likely that a combination of approaches could yield better results than either alone. We therefore integrate a data-driven classifier for the special task of PP attachment into an existing rule-based parser and measure the effect that the additional information has on the overall accuracy.

2 Motivation

PP attachment disambiguation has often been studied as a benchmark test for empirical methods in natural language processing. Prepositions allow subordination to many different attachment sites, and the choice between them is influenced by factors from many different linguistic levels, which are generally subject to preferential rather than rigorous regularities. For this reason, PP attachment is a comparatively difficult subtask for rule-based syntax analysis and has often been attacked by statistical methods.

Because probabilistic approaches solve PP attachment as a natural subtask of parsing anyhow, the obvious application of a PP attacher is to integrate it into a rule-based system. Perhaps surprisingly, so far this has rarely been done. One reason for this is that many rule-driven syntax analyzers provide no obvious way to integrate uncertain, statistical information into their decisions. Another is the traditional emphasis on PP attachment as a binary classification task; since (Hindle and Rooth, 1991), research has concentrated on resolving the ambiguity in the category pattern 'V+N+P+N', i.e. predicting the PP attachment to either the verb or the first noun. It is often assumed that the correct attachment is always among these

two options, so that all problem instances can be solved correctly despite the simplification. This task is sufficient to measure the relative quality of different probability models, but it is quite different from what a parser must actually do: It is easier because the set of possible answers is pre-filtered so that only a binary decision remains, and the baseline performance for pure guessing is already 50%. But it is harder because it does not provide the predictor with all the information needed to solve many doubtful cases; (Hindle and Rooth, 1991) found that human arbiters consistently reach a higher agreement when they are given the entire sentence rather than just the four words concerned.

Instead of the accuracy of PP attachers in the *isolated* decision between two words, we investigate the problem of *situated* PP attachment. In this task, all nouns and verbs in a sentence are potential attachment points for a preposition; the computer must find suitable attachments for one or more prepositions in parallel, while building a globally coherent syntax structure at the same time.

3 Methods

Statistical PP attachment is based on the observation that the identities of content words can be used to predict which prepositional phrases modify which words, and achieve better-than-chance accuracy. This is apparently because, as heads of their respective phrases, they are representative enough that they can serve as a crude approximation of the semantic structure that could be derived from the phrases. Consider the following example (the last sentence in our test set):

Die Firmen müssen noch die Bedenken der EU-Kommission gegen die Fusion ausräumen. (The companies have yet to address the Commission's concerns about the merger.)

In this sentence, the preferred analysis will pair the preposition 'gegen' *(against, about, versus)* with the noun 'Bedenken' *(concerns)*, since the proposition is clearly that the concerns pertain to the merger. A syntax tree of this interpretation is shown in Figure 1. Note that there are at least three different syntactically plausible attachment sites for the preposition. In fact, there are even more, since a parser can make no initial assumptions about the global structure of the syntax tree that it will construct; for instance, the possibility that 'gegen' attaches to the noun 'Firmen' *(companies)* cannot be ruled out when beginning to parse.

3.1 WCDG

For the following experiments, we used the dependency parser of German described in (Foth et al., 2005). This system is especially suited to our goals for several reasons. Firstly, the parser achieves the highest published dependency-based accuracy on unrestricted written German input, but still has a comparatively high error rate for prepositions. In particular, it mis-attaches the preposition 'gegen' in the example sentence. Second, although rule-based in nature, it uses numerical penalties to arbitrate between different disambiguation rules. It is therefore easy to add another rule of varying strength, which depends on the output of an external statistical predictor, to guide the parser when it has no other means of making an attachment decision. Finally, the parser and grammar are freely available for use and modification (`http://nats-www.informatik.uni-hamburg.de/download`).

Weighted Constraint Dependency Grammar (Schröder, 2002) models syntax structure as labelled dependency trees as shown in the example. A grammar in this formalism is written as a set of *constraints* that license well-formed partial syntax structures. For instance, general projectivity rules ensure that the dependency tree corresponds to a properly nested syntax structure without crossing brackets[1]. Other constraints require an auxiliary verb to be modified by a full verb, or prescribe morphosyntactical agreement between a determiner and its regent (the word modified by the determiner). Although the *Constraint Satisfaction Problem* that this formalism defines is, in theory, infeasibly hard, it can nevertheless be solved approximatively with heuristic solution methods, and achieve competitive parsing accuracy.

To allow the resolution of true ambiguity (the existence of different structures neither of which is strictly ungrammatical), *weighted* constraints can be written that the solution *should* satisfy, if this is possible. The goal is then to build the structure that violates as few constraints as possible, and preferentially violates weak rather than strong constraints. This allows preferences to be expressed rather than hard rules. For instance, agreement constraints could actually be declared as violable, since typing errors, reformulations, etc. can

[1] Some constructions of German actually violate this property; exceptions in the projectivity constraints deal with these cases.

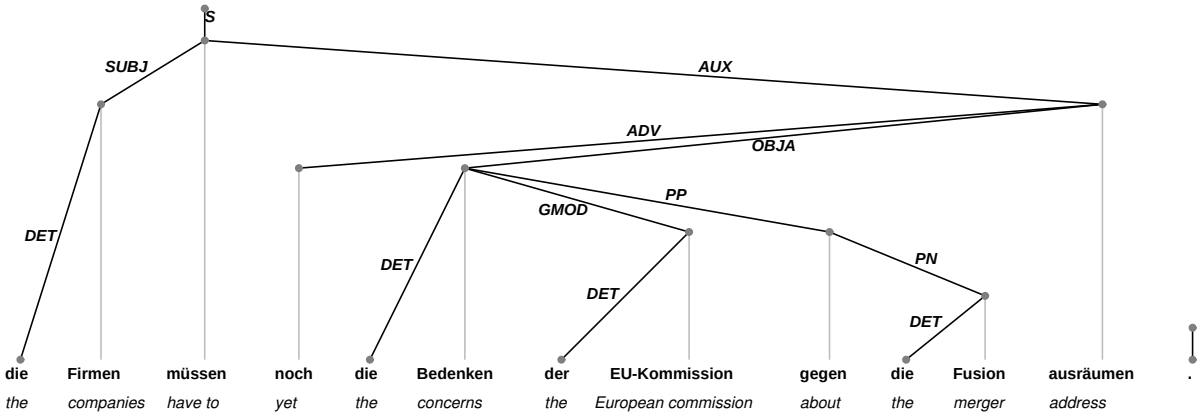

Figure 1: Correct syntax analysis of the example sentence.

and do actually lead to mis-inflected phrases. In this way robustness against many types of error can be achieved while still preferring the correct variant. For more about the WCDG parser, see (Schröder, 2002; Foth and Menzel, 2006) .

The grammar of German available for this parser relies heavily on weighted constraints both to cope with many kinds of imperfect input and to resolve true ambiguities. For the example sentence, it retrieves the desired dependencies except for constructing the implausible dependency 'ausräumen'+'gegen' *(address against)*. Let us briefly review the relevant constraints that cause this error:

- General structural, valence and agreement constraints determine the macro structure of the sentence in the desired way. For instance, the finite and the full verb must combine to form an auxiliary phrase, because this is the only way of accounting for all words while satisfying valence and category constraints. For the same reasons both determiners must be paired with their respective nouns. Also, the prepositional phrase itself is correctly predicted.

- General category constraints ensure that the preposition can attach to nouns and verbs, but not, say, to a determiner or to punctuation.

- A weak constraint on adjuncts says that adjuncts are usually close to their regent. The penalty of this constraint varies according to the length of the dependency that it is applied to, so that shorter dependencies are generally preferred.

- A slightly stronger constraint prefers attachment of the preposition to the verb, since overall verb attachment is more common than noun attachment in German. Therefore, the verb attachment leads to the globally best solution for this sentence.

There are no lexicalized rules that capture the particular plausibility of the phrase 'Bedenken gegen' *(concerns about)*. A constraint that describes this individual word pair would be trivial to write, but it is not feasible to model the general phenomenon in this way; thousands of constraints would be needed just to reflect the more important collocations in a language, and the exact set of collocating words is impossible to predict accurately. Data-driven information would be much more suitable for curing this lexical blind spot.

3.2 The Collocation Measure

The usual way to retrieve the lexical preference of a word such as 'Bedenken' for 'gegen' is to obtain a large corpus and assume that it is representative of the entire language; in particular, that collocations in this corpus are representative of collocations that will be encountered in future input. The assumption is of course not entirely true, but it can nevertheless be preferable to rely on such uncertain knowledge rather than remain undecided, on the reasonable assumption that it will lead to more correct than wrong decisions. Note that the same reasoning applies to many of the violable constraints in a WCDG: although they do not hold on *all* possible structures, they hold more often than they fail, and therefore can be useful for analysing unknown input.

Different measures have been used to gauge the strength of a lexical preference, but in general the efficacy of the statistical approach depends more on the suitability of the training corpus than on details of the collocation measure. Since our focus

is not on finding the best extraction method, but on judging the benefit of statistical components to parsing, we employ a collocation measure related to the idea of *mutual information*: a collocation between a word w and a preposition p is judged more likely the more often it appears, and the less often its component words appear. By normalizing against the total number t of utterances we derive a measure of Lexical Attraction for each possible collocation:

$$LA(w,p) := \frac{f_{w+p}}{t} \Big/ \left(\frac{f_w}{t} \cdot \frac{f_p}{t} \right)$$

For instance, if we assume that the word 'Bedenken' occurs in one out of 2,000 sentences of German and the word 'gegen' occurs in one sentence out of 31 (these figures were taken from the unsupervised experiment described later), then pure chance would make the two words co-occur in one sentence out of 62,000. If the LA score is higher than 1, i. e. we observe a much higher frequency of co-occurrences in a large corpus, we can assume that the two events are not statistically independent — in other words, that there is a positive correlation between the two words. Conversely, we would expect a much lower score for the implausible collocation 'Bedenken'+'für', indicating a dispreference for this attachment.

4 Experiments

4.1 Sources

To obtain the counts to base our estimates of attraction on, we first turned to the dependency treebank that accompanies the WCDG parsing suite. This corpus contains some 59,000 sentences with 1,000,000 words with complete syntactic annotations, 61% of which are drawn from online technical newscasts, 33% from literature and 6% from law texts. We used the entire corpus except for the test set as a source for counting PP attachments directly. All verbs, nouns and prepositions were first reduced to their base forms in order to reduce the parameter space. Compound nouns were reduced to their base nouns, so that 'EU-Kommission' is treated the same as 'Kommission', on the assumption that the compound exerts similar attractions as the base noun. In contrast, German verbs with prefixes usually differ markedly in their preferences from the base verb. Since forms of verbs such as 'ausräumen' *(address)* can be split into two parts ('NP räumte NP aus'), such separated verbs were reassembled before stemming.

(w,p)	f_{w+p}	f_w	LA
'Firma'+'gegen'	72	76492	0.03
'Bedenken'+'gegen'	1529	9618	4.96
'Kommission'+'gegen'	223	52415	0.13
'ausräumen'+'gegen'	130	2342	1.73
(where $f_p = 566068$, $t = 17657329$)			

Table 1: Example calculation of lexical attraction.

Although the information retrieved from complete syntax trees is valuable, it is clearly insufficient for estimating many valid collocations. In particular, even for a comparatively strong collocation such as 'Bedenken'+'gegen' we can expect only very few instances. (There are, in fact, 4 such instances, well above chance level but still a very small number.) Therefore we used the archived text from 18 volumes of the newspaper *tageszeitung* as a second source. This corpus contains about 295,000,000 words and should allow us to detect many more collocations. In fact, we do find 2338 instances of 'Bedenken'+'gegen' in the same sentence.

Of course, since we have no syntactic annotations for this corpus (and it would be infeasible to create them even by fully automatic parsing), not all of these instances may indicate a syntactic dependency. (Ratnaparkhi, 1998) solved this problem by regarding only prepositions in syntactically unambiguous configurations. Unfortunately, his patterns cannot directly be applied to German sentences because of their freer word order. As an approximation it would be possible to count only pairs of adjacent content words and prepositions. However, this would introduce systematic biases into the counts, because nouns do in fact very often occur adjacently to prepositions that modify them, but many verbs do not. For instance, the phrase 'jmd. anklagen wegen etw.' *(to sue s.o. for s.th.)* gives rise to a strong collocation between the verb 'anklagen' and the preposition 'wegen'; however, in the predominant sentence types of German, the two words are virtually never adjacent, because either the preposition kernel or the direct object must intervene. Therefore, we relax the adjacency condition for verb attachment and also count prepositions that occur within a fixed distance of their suspected regent.

Table 1 shows the detailed values when judging the example sentence according to the unparsed corpus. The strong collocation that we would expect for 'Bedenken'+'gegen' is indeed

Value of i	Recall for V	for N	overall
1	96.2%	39.8%	65.2%
2	96.2%	52.0%	71.9%
5	88.8%	66.3%	76.4%
8	80.0%	79.6%	79.8%
10	67.5%	82.7%	75.8%

Table 2: Influence of noun factor on solving isolated attachment decisions.

observed, with a value of 4.96. However, the verb attachment also has a score above 1, indicating that 'gegen'+'ausräumen' *(to address about)* are also positively correlated. This is almost certainly a misleading figure, since those two words do not form a plausible verb phrase; it is much more probable that the very strong, in fact idiomatic, correlation 'Bedenken ausräumen' *(to address concerns)* causes many co-occurrences of all *three* words. Therefore our figures falsely suggest that 'gegen' would often attach to 'ausräumen', when it is in fact the direct object of that verb that it is attracted to.

(Volk, 2002) already suggested that this counting method introduced a general bias toward verb attachment, and when comparing the results for very frequent words (for which more reliable evidence is available from the treebank) we find that verb attachments are in fact systematically overestimated. We therefore adopted his approach and artificially inflated all noun+preposition counts by a constant factor i. To estimate an appropriate value for this factor, we extracted 178 instances of the standard verb+noun+preposition configuration from our corpus, of which 80 were verb attachments (V) and 98 were noun attachments (N).

Table 2 shows the performance of the predictor for this binary decision task. Taken as it is, it retrieves most verb attachments, but less than half of the noun attachments, while higher values of i can improve the recall both for noun attachments and overall. The performance achieved falls somewhat short of the highest figures reported previously for PP attachment for German (Volk, 2002); this is at least in part due to our simple model that ignores the kernel noun of the PP. However, it could well be good enough to be integrated into a full parser and provide a benefit to it. Also, the syntactical configuration in this standard benchmark is not the predominant one in complete German sentences; in fact fewer than 10% of all prepositions occur in this context. The best performance on the triple task is therefore not guaranteed to be the best choice for full parsing. In our experiments, we used a value of $i = 8$, which seems to be suited best to our grammar.

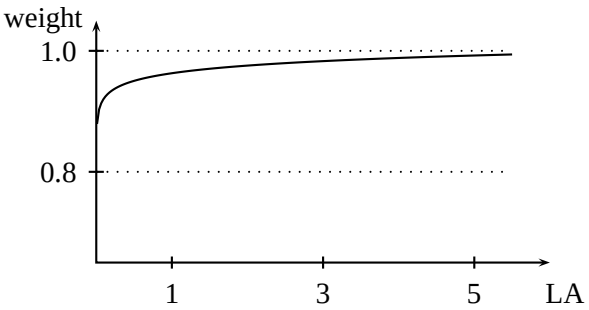

Figure 2: Mapping lexical attraction values to penalties

4.2 Integration Method

To add our simple collocation model to the parser, it is sufficient to write a single variable-strength constraint that judges each PP dependency by how strong the lexical attraction between the regent and the dependent is. The only question is how to map our lexical attraction values to penalties for this constraint. Their predicted relative order of plausibility should of course be reflected, so that dependencies with a high lexical attraction are preferred over those with lower lexical attraction. At the same time, the information should not be given too much weight compared to the existing grammar rules, since it is heuristic in nature and should certainly not override important principles such as valence or agreement. The penalties of WCDG constraints range from 0.0 (hard constraint) through 1.0 (a constraint with this penalty has no effect whatsoever and is only useful for debugging).

We chose an inverse mapping based on the logarithm of lexical attraction (cf. Figure 2):

$$p(w,p) = \frac{\max(1,\min(0.8,1-(2-\log_3(LA(w,p)))/50))}{\mu}$$

where μ is a normalization constant that scales the highest occurring value of LA to 1. For instance, this mapping will interpret a strong lexical attraction of 5 as the penalty 0.989 (almost perfect) and a lexical attraction of only 0.5 as the penalty 0.95 (somewhat dispreferred). The overall range of PP attachment penalties is limited to the interval $[0.8 - 1.0]$, which ensures that the judgement of the statistical module will usually come into play only when no other evidence is available; preliminary experiments showed that a stronger integration of the component yields no additional advantage. In any case, the exact figure depends closely on the valuation of the existing constraints of the grammar and is of little importance as such.

Label	occurred	retrieved	errors	accuracy
PP	1892	1285	607	67.9%
ADV	1137	951	186	83.6%
OBJA	775	675	100	87.1%
APP	659	567	92	86.0%
SUBJ	1338	1251	87	93.5%
S	1098	1022	76	93.1%
KON	481	406	75	84.4%
REL	167	107	60	64.1%
overall	17719	16073	1646	90.7

Table 3: Performance of the original parser on the test set.

Besides adding the new constraint 'PP attachment' to the grammar, we also disabled several of the existing constraints that apply to prepositions, since we assume that our lexicalized model is superior to the unlexicalized assumptions that the grammar writers had made so far. For instance, the constraint mentioned in Section 3 that globally prefers verb attachment to noun attachment is essentially a crude approximation of lexical attraction, whose task is now taken over entirely by the statistical predictor. We also assume that lexical preference exerts a stronger influence on attachment than mere linear distance; therefore we changed the distance constraint so that it exempts prepositions from the normal distance penalties imposed on adjuncts.

4.3 Corpus

For our parsing experiments, we used the first 1,000 sentences of technical newscasts from the dependency treebank mentioned above. This test set has an average sentence length of 17.7 words, and from previous experiments we estimate that it is comparable in difficulty to the NEGRA corpus to within 1% of accuracy. Although online articles and newspaper copy follow some different conventions, we assume the two text types are similar enough that collocations extracted from one can be used to predict attachments in the other.

For parsing we used the heuristic transformation-based search described in (Foth et al., 2000). Table 3 illustrates the structural accuracy[2] of the unmodified system for various subordination types. For instance, of the 1892 dependency edges with the label 'PP' in the gold standard, 1285 are attached correctly by the parser, while 607 receive an incorrect regent. We see that PP attachment decisions are particularly prone to errors both in absolute and in relative terms.

[2]Note that the WCDG parser always succeeds in assigning exactly one regent to each word, so that there is no difference between precision and recall. We refer to structural accuracy as the ratio of words which have been attached correctly to all words.

Method	PP accuracy	overall accuracy
baseline	67.9%	90.7%
supervised	79.4%	91.9%
unsupervised	78.3%	91.9%
backed-off	78.9%	92.2%

Table 4: Structural accuracy of PP edges and all edges.

4.4 Results

We trained the PP attachment predictor both with the counts acquired from the dependency treebank (supervised) and those from the newspaper corpus (unsupervised). We also tested a mode of operation that uses the more reliable data from the treebank, but backs off to unsupervised counts if the hypothetical regent was seen fewer than 1,000 times in training.

Table 4 shows the results when parsing with the augmented grammar. Both the overall structural accuracy and the accuracy of PP edges are given; note that these figures result from the general subordination task, therefore they correspond to Table 3 and not to Table 2. As expected, lexicalized preference information for prepositions yields a large benefit to full parsing: the attachment error rate is decreased by 34% for prepositions, and by 14% overall. In this experiment, where much more unsupervised training data was available, supervised and unsupervised training achieved almost the same level of performance (although many individual sentences were parsed differently).

A particular concern with corpus-based decision methods is their applicability beyond the training corpus. In our case, the majority of the material for supervised training was taken from the same newscast collection as the test set. However, comparable results are also achieved when applying the parser to the standard test set from the NEGRA corpus of German, as used by (Schiehlen, 2004; Foth et al., 2005): adding the PP predictor trained on our dependency treebank raises the overall attachment accuracy from 89.3% to 90.6%. This successful reuse indicates that lexical preference between prepositions and function words is largely independent of text type.

5 Related Work

(Hindle and Rooth, 1991) first proposed solving the prepositional attachment task with the help of statistical information, and also defined the prevalent formulation as a binary decision problem with three words involved. (Ratnaparkhi et al., 1994)

extended the problem instances to quadruples by also considering the kernel noun of the PP, and used maximum entropy models to estimate the preferences.

Both supervised and unsupervised training procedures for PP attachment have been investigated and compared in a number of studies, with supervised methods usually being slightly superior (Ratnaparkhi, 1998; Pantel and Lin, 2000), with the notable exception of (Volk, 2002), who obtained a worse accuracy in the supervised case, obviously caused by the limited size of the available treebank. Combining both methods can lead to a further improvement (Volk, 2002; Kokkinakis, 2000), a finding confirmed by our experiments.

Supervised training methods already applied to PP attachment range from stochastic maximum likelihood (Collins and Brooks, 1995) or maximum entropy models (Ratnaparkhi et al., 1994) to the induction of transformation rules (Brill and Resnik, 1994), decision trees (Stetina and Nagao, 1997) and connectionist models (Sopena et al., 1998). The state-of-the-art is set by (Stetina and Nagao, 1997) who generalize corpus observations to semantically similar words as they can be derived from the WordNet hierarchy.

The best result for German achieved so far is the accuracy of 80.89% obtained by (Volk, 2002). Note, however, that our goal was not to optimize the performance of PP attachment in isolation but to quantify the contribution it can make to the performance of a full parser for unrestricted text.

The accuracy of PP attachment has rarely been evaluated as a subtask of full parsing. (Merlo et al., 1997) evaluate the attachment of multiple prepositions in the same sentence for English; 85.3% accuracy is achieved for the first PP, 69.6% for the second and 43.6% for the third. This is still rather different from our setup, where PP attachment is fully integrated into the parsing problem. Closer to our evaluation scenario comes (Collins, 1999) who reports 82.3%/81.51% recall/precision on PP modifications for his lexicalized stochastic parser of English. However, no analysis has been carried out to determine which model components contributed to this result.

A more application-oriented view has been adopted by (Schwartz et al., 2003), who devised an unsupervised method to extract positive and negative lexical evidence for attachment preferences in English from a bilingual, aligned English-Japanese corpus. They used this information to reattach PPs in a machine translation system, reporting an improvement in translation quality when translating into Japanese (where PP attachment is not ambiguous and therefore matters) and a decrease when translating into Spanish (where attachment ambiguities are close to the original ones and therefore need not be resolved).

Parsing results for German have been published a number of times. Combining treebank transformation techniques with a suffix analysis, (Dubey, 2005) trained a probabilistic parser and reached a labelled F-score of 76.3% on phrase structure annotations for a subset of the sentences used here (with a maximum length of 40). For dependency parsing a labelled accuracy of 87.34% and an unlabelled one of 90.38% has been achieved by applying the dependency parser described in (McDonald et al., 2005) to German data. This system is based on a procedure for online large margin learning and considers a huge number of locally available features, which allows it to determine the optimal attachment fully deterministically. Using a stochastic variant of Constraint Dependency Grammar (Wang and Harper, 2004) reached a 92.4% labelled F-score on the Penn Treebank, which slightly outperforms (Collins, 1999) who reports 92.0% on dependency structures automatically derived from phrase structure results.

6 Conclusions and future work

Corpus-based data has been shown to provide a significant benefit when used to guide a rule-based dependency parser of German, reducing the error rate for situated PP attachment by one third. Prepositions still remain the largest source of attachment errors; many reasons can be tracked down for individual errors, such as faulty POS tagging, misinterpreted global sentence structure, genuinely ambiguous constructions, failure of the attraction heuristics, or simply lack of processing time. However, considering that even human arbiters often agree only on 90% of PP attachments, the results appear promising. In particular, many attachment errors that strongly disagree with human intuition (such as in the example sentence) were in fact prevented. Thus, the addition of a corpus-based knowledge source to the system yielded a much greater benefit than could have been achieved with the same effort by writing individual constraints.

One obvious further task is to improve our simple-minded model of lexical attraction. For instance, some remaining errors suggest that taking the kernel noun into account would yield a higher attachment precision; this will require a redesign of the extraction tools to keep the parameter space manageable. Also, other subordination types than 'PP' may benefit from similar knowledge; e.g., in many German sentences the roles of subject and object are syntactically ambiguous and can only be understood correctly through world knowledge. This is another area in which synergy between lexical attraction estimates and general symbolic rules appears possible.

References

E. Brill and P. Resnik. 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proc. 15th Int. Conf. on Computational Linguistics*, pages 1198 – 1204, Kyoto, Japan.

M. Collins and J. Brooks. 1995. Prepositional attachment through a backed-off model. In *Proc. of the 3rd Workshop on Very Large Corpora*, pages 27–38, Somerset, New Jersey.

M. Collins. 1999. *Head-Driven Statistical Models for Natural Language Parsing*. Phd thesis, University of Pennsylvania, Philadephia, PA.

A. Dubey. 2005. What to do when lexicalization fails: parsing German with suffix analysis and smoothing. In *Proc. 43rd Annual Meeting of the ACL*, Ann Arbor, MI.

K. Foth and W. Menzel. 2006. Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In *Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006*, Sydney.

K. Foth, W. Menzel, and I. Schröder. 2000. A Transformation-based Parsing Technique with Anytime Properties. In *4th Int. Workshop on Parsing Technologies, IWPT-2000*, pages 89 – 100.

K. Foth, M. Daum, and W. Menzel. 2005. Parsing unrestricted German text with defeasible constraints. In H. Christiansen, P. R. Skadhauge, and J. Villadsen, editors, *Constraint Solving and Language Processing*, volume 3438 of *LNAI*, pages 140–157. Springer-Verlag, Berlin.

D. Hindle and M. Rooth. 1991. Structural Ambiguity and Lexical Relations. In *Meeting of the Association for Computational Linguistics*, pages 229–236.

D. Kokkinakis. 2000. Supervised pp-attachment disambiguation for swedish; (combining unsupervised supervised training data). *Nordic Journal of Linguistics*, 3.

R. McDonald, F. Pereira, K. Ribarov, and J. Hajic. 2005. Non-projective dependency parsing using spanning tree algorithms. In *Proc. Human Language Technology Conference / Conference on Empirical Methods in Natural Language Processing, HLT/EMNLP-2005*, Vancouver, B.C.

P. Merlo, M. Crocker, and C. Berthouzoz. 1997. Attaching Multiple Prepositional Phrases: Generalized Backed-off Estimation. In *Proc. 2nd Conf. on Empirical Methods in NLP*, pages 149–155, Providence, R.I.

P. Pantel and D. Lin. 2000. An unsupervised approach to prepositional phrase attachment using contextually similar words. In *Proc. 38th Meeting of the ACL*, pages 101–108, Hong Kong.

A. Ratnaparkhi, J. Reynar, and S. Roukos. 1994. A Maximum Entropy Model for Prepositional Phrase Attachment. In *Proc. ARPA Workshop on Human Language Technology*, pages 250 –255.

A. Ratnaparkhi. 1998. Statistical models for unsupervised prepositional phrase attachment. In *Proc. 17th Int. Conf. on Computational Linguistics*, pages 1079–1085, Montreal.

M. Schiehlen. 2004. Annotation Strategies for Probabilistic Parsing in German. In *Proceedings of COLING 2004*, pages 390–396, Geneva, Switzerland, Aug 23–Aug 27. COLING.

I. Schröder. 2002. *Natural Language Parsing with Graded Constraints*. Ph.D. thesis, Department of Informatics, Hamburg University, Hamburg, Germany.

L. Schwartz, T. Aikawa, and C. Quirk. 2003. Disambiguation of english PP-attachment using multilingual aligned data. In *Machine Translation Summit IX*, New Orleans, Louisiana, USA.

J. M. Sopena, A. LLoberas, and J. L. Moliner. 1998. A connectionist approach to prepositional phrase attachment for real world texts. In *Proc. 17th Int. Conf. on Computational Linguistics*, pages 1233–1237, Montreal.

J. Stetina and M. Nagao. 1997. Corpus based PP attachment ambiguity resolution with a semantic dictionary. In Jou Shou and Kenneth Church, editors, *Proc. 5th Workshop on Very Large Corpora*, pages 66–80, Hong Kong.

M. Volk. 2002. Combining Unsupervised and Supervised Methods for PP Attachment Disambiguation. In *Proc. of COLING-2002*, Taipeh.

W. Wang and M. P. Harper. 2004. A statistical constraint dependency grammar (CDG) parser. In *Proc. ACL Workshop Incremental Parsing: Bringing Engineering and Cognition Together*, pages 42–49, Barcelona, Spain.