Getting a Grip on Morphological Disambiguation

Erhard W. Hinrichs and Julia S. Trushkina
Eberhard-Karls-Universität Tübingen
Seminar für Sprachwissenschaft, Computerlinguistik
Wilhelmstr. 113
D-72074 Tübingen
email: {eh,jul}@sfs.uni-tuebingen.de

Abstract

The paper argues that morphological disambiguation is a crucial step for assignment of dependency structures. Quantitative evaluation on a German corpus shows that morphological disambiguation of NPs together with syntactic heuristics yields unique morphological analyses for the assignment of dependency relations to German NPs in 77.08% of all cases.

1 Introduction

The research reported here is part of a larger project on the development of a robust parsing scheme GRIP (GeRman Incremental Parsing) that uses the Xerox Incremental Deep Parsing System (XIP) (Ait and Chanod and Roux, to appear) and provides syntactic annotation in an incremental fashion: after textual input is tokenized, morphologically analyzed and disambiguated, syntactic annotation is added in two distinct stages of processing. First, a chunk parser provides a partial constituent analysis. In a second stage, the chunked input is further annotated by dependency links that reflect the function-argument structure for each chunked clause. This latter stage of processing is inspired by ideas originating in frameworks of dependency grammar which express grammatical relations as independent notions, rather than as a secondary concept derivable from constituent structure only.[1]

The current paper addresses one specific subtask in the overall GRIP parsing scheme: morphological disambiguation. We will demonstrate that morphological disambiguation is a crucial step in narrowing down the search space for the correct assignment of dependency structures, particularly for languages with rich inflectional morphology. Furthermore, we will describe in detail the customized disambiguation rules of XIP that provide the necessary computational tools to efficiently carry out morphological disambiguation.

The importance of morphological disambiguation has been recognized by a number of researchers, in particular to improve the accuracy of morphological analysis (Oflazer, 1997) and of part-of-speech tagging (Hajic, 2001), (Voutilainen, 1995). We will compare our approach to this previous body of research in detail in section 4.

2 Incremental Syntactic Annotation

Due to its incremental nature, GRIP crucially relies on the accuracy of annotation at previous levels. Chunking will depend on the accuracy of part-of-speech disambiguation, while dependency parsing relies crucially on the structure of the pre-chunked input and on the morphological properties of individual chunks. For example, in order to determine the subject of a clause, case and number information associated with the NP chunks that occur in the clause is of primary importance. For languages with rich inflectional morphology, it can often be difficult to determine such case and number information uniquely since one and the same word form may be associated with more than one combination of case, number and gender values. Consider the German sentence in (1):

(1) Die Politiker gaben verdienten
the politicians gave worthy
Beamten und Lohnempfängern ein
civil servants and wage recipients a
höheres Gehalt.
higher salary

[1] For recent applications of dependency grammars to syntactic annotation and parsing see, among others, Tapanainen and Järvinen (1994) and Duchier (1999).

'The politicians gave worthy civil servants and wage recipients a higher salary.'

The only morphologically unambiguous noun phrase in (1) is the NP *Lohnempfängern* with the unique analysis `Noun+Masc+Pl+Dat`. As shown in (2)–(4), the lexical nodes for all other NPs in the sentence are morphologically many times ambiguous. The analyses are provided by the morphological analyzer for German developed by the Xerox Research Centre Europe (XRCE).[2]

(2)

Die	Pron+Dem+FMN+Pl+NomAcc
Die	Pron+Dem+Fem+Sg+NomAcc
Die	Pron+Rel+FMN+Pl+NomAcc
Die	Pron+Rel+Fem+Sg+NomAcc
Die	Det+Def+Fem+Sg+NomAcc+St
Die	Det+Def+FMN+Pl+NomAcc+St
Politiker	Noun+Masc+Sg+NomAccDat
Politiker	Noun+Masc+Pl+NomAccGen

For example, the noun *Politiker* has a unique value only for gender (`Masc`). Number and case values are not unique and co-vary.[3] The preceding token *die* exhibits a three-way word class ambiguity between a determiner reading (`Det`), a demonstrative pronoun reading, and a relative pronoun reading. The latter two will in all likelihood be eliminated by a reliable part-of-speech tagger. However, even for the remaining determiner reading there are several distinct readings: *die*, taken in isolation, can be (i) nominative or accusative, singular, feminine, or (ii) nominative or accusative plural for any gender.[4] However, in the context of the following noun *Politiker*, only the latter reading is valid since it matches the gender specification of the noun. In the other direction, the determiner also helps to partially disambiguate the contextually valid readings of the noun by retaining as possible values of case nominative and accusative. The discussion of this first example shows the nature of this kind of contextual morphological disambiguation: lexical nodes within the same NP mutually constrain each other as to the set of possible readings.

While example (2) requires identity of case, number and gender values between determiner and noun, other combinations of lexical categories require distinct values for certain morphological features.[5] In German, word forms for adjectives and determiners can be classified as belonging to either weak or strong declension classes.[6] For example, all forms of the definite determiner *der* belong to the strong declension class, while the paradigm of the indefinite determiner *ein* is split between weak and strong forms. In addition, some nouns, in particular those derived from adjectives like *Beamter*, also exhibit a distinction between weak and strong forms.

If determiners co-occur with adjectives and nouns in the same NP, adjective and noun agree in declension class, whereas the declension value of the determiner is the opposite. The NP *ein höheres Gehalt* and the set of candidate analyses in (3) demonstrate this. The only contextually valid reading is the sequence `Det+Indef+Neut+Sg+NomAcc+Wk`, `Adj+Neut+Sg+NomAcc+St`, `Noun+Neut+Sg+NomAcc`.

(3)

ein	Det+Indef+Masc+Sg+Nom+Wk
ein	Det+Indef+Neut+Sg+NomAcc+Wk
höheres	Adj+Neut+Sg+NomAcc+St
Gehalt	Noun+Neut+Sg+NomAccDat
Gehalt	Noun+Masc+Sg+NomAccDat

The morphological analysis for the NP in (4) exemplifies agreement of declension values between adjective and noun. In this case the set of contextually valid readings is still quite large since all adjectival readings that are compatible with the gender specification of the noun will be

[2] An on-line demo version of the XRCE morphological analyzer is available at `www.xrce.xerox.com/competencies/content-analysis/demos/german.de.html`.

[3] The morphological tag `NomAccGen` stands for *nominative*, *accusative* or *genitive*.

[4] The morphological tag `FMN` stands for any gender.

[5] Even for determiners and nouns, identity of case, number and gender values is sometimes too strong a constraint. If the determiner is realized by a relative pronoun as in *dessen Mutter* ('whose mother'), a mismatch in case values needs to be allowed.

[6] For a comprehensive study of the distributional properties of weak and strong forms in German NPs see Zwicky (1986).

retained. However, further pruning of contextually valid readings is possible. If noun phrases do not include an overt determiner, as in (4), then only strong forms are grammatical, and all weak forms can be eliminated. Furthermore, in example sentence (1), the NP in (4) is coordinated with the NP *Lohnempfängern*, which is unambiguously `Noun+Masc+Pl+Dat`. Since conjoined NPs have to agree in case, the noun in (4) also has to be `Dat`. Thus, the only contextually valid reading for the NP in (4) is the sequence of morphological tags `Adj+Masc+Pl+Dat+St`, `Noun+Masc+Pl+Dat+St`.[7]

(4)

verdienten	Adj+Fem+Sg+DatGen+Wk
verdienten	Adj+Masc+Sg+AccGen+StWk
verdienten	Adj+Masc+Sg+Dat+Wk
verdienten	Adj+Neut+Sg+Gen+StWk
verdienten	Adj+Neut+Sg+Dat+Wk
verdienten	Adj+FMN+Pl+NomAccDatGen+Wk
verdienten	Adj+FMN+Pl+Dat+St
Beamten	Noun+Masc+Sg+AccGen+StWk
Beamten	Noun+Masc+Sg+Dat+Wk
Beamten	Noun+Masc+Pl+NomAccDatGen+Wk
Beamten	Noun+Masc+Pl+Dat+St

3 Guiding Dependency Parsing by Morphological Disambiguation

The preceding discussion was designed to provide an overview of some of the empirical issues involved in morphological disambiguation for a morphologically rich language like German. The present section will demonstrate the utility of morphological disambiguation for further incremental syntactic annotation.

Consider once again our example sentence in (1). The ultimate goal for syntactic annotation with the XIP System is to assign a dependency structure to the input sentence. As an intermediate stage, the input is chunked into major constituents. This chunked structure then serves as input to the dependency analysis.

The intended dependency structure output based on the chunked structure is shown in (5).[8]

(5) {VF {NP#1 Die Politiker}} {LK#2 gaben} {MF {NP#3 verdienten Beamten} und {NP#4 Lohnempfängern} {NP#5 ein höheres Gehalt}}.

SUBJ(#2,#1), OBJ_dir(#2,#5),
OBJ_indir(#2,#3), OBJ_indir(#2,#4)

In GRIP, the dependency analysis is constructed with the aid of lexical resources such as CELEX and IMS-LEX which provide subcategorization information for German verbs. A simplified entry of the information that CELEX provides for the lemma *geben* is shown in (6).

(6) geben +VERB+Aux_H+Acc_Comp+
+Dat_Comp+Comp_Subj

The key to identifying the correct dependency links is to try to match the case specifications inherent in the chunk analysis with the subcategorization information provided by the lexical entry of the main verb. Here is where morphological disambiguation plays a crucial role. In the previous section, we discussed how co-occurring lexical nodes mutually constrain the set of contextually valid morphological interpretations. For the four NPs in sentence (1), the set of analyses shown in (7) will be retained.

The remaining ambiguities concern the case values of the NPs *verdienten Beamten*, *die Politiker* and *ein höheres Gehalt*. As discussed before, *verdienten Beamten* is coordinated with the NP *Lohnempfängern*; thus, the former is unambiguously `Noun+Masc+Pl+Dat`. The NPs *die Politiker* and *ein höheres Gehalt* can be either nominative or accusative case. Thus, in principle, both NPs can serve as either the subject or direct object of the finite verb. However, this ambiguity can be resolved due to subject-verb agreement. Since the finite verb is plural, only the plural NP *die Politiker* can be the subject, and the NP *ein höheres Gehalt* should be the direct object.

[7] The morphological tag `FMN`, which stands for any gender, is disambiguated for the adjective as `Masc` due to the gender specification on the noun.

[8] Apart from NP chunking, the structure in (5) labels the topological fields of the clause. Such a topological field analysis is useful for identifying the overall structure of the clause (see Hinrichs et al. (2000) for details).

The node structure in (5) is grossly oversimplified: it leaves out morphological information percolated up from the daughter nodes. How such morphological information is percolated to the phrasal nodes will be explained in detail in section 5.

(7)

Die	Det+Def+Masc+Pl+NomAcc+St
Politiker	Noun+Masc+Pl+NomAcc
verdienten	Adj+Masc+Pl+Dat+St Adj+Masc+Sg+AccGen+St
Beamten	Noun+Masc+Pl+Dat+St Noun+Masc+Sg+AccGen+St
Lohnempfängern	Noun+Masc+Pl+Dat
ein	Det+Indef+Neut+Sg+NomAcc+Wk
höheres	Adj+Neut+Sg+NomAcc+St
Gehalt	Noun+Neut+Sg+NomAcc

What this example has shown is that morphological disambiguation in conjunction with other morpho-syntactic constraints such as subject-verb agreement can effectively reduce the number of candidate readings and uniquely determine the dependency structure to be assigned. In the remainder of the paper we will discuss how the XIP System provides the necessary computational environment to efficiently carrying out morphological disambiguation.

4 Implementing Morphological Disambiguation

XIP provides two types of disambiguation rules: ordinary disambiguation rules (ODRs), which can eliminate readings for a single lexical node on the basis of left and/or right contexts of the token, and double reduction rules (DRRs), which simultaneously reduce readings of sequences of tokens. The entire rule set is organized by levels, which determines the order of application. Each level may consist of one or more rules of one kind.[9] However, it is not required that all ODRs precede all DRRs, or vice versa.

The general format for ODRs is shown in (8).

(8) readings_filter = |left_context| selected_readings |right_context|.

The left side of the rule specifies to which readings of lexical nodes the disambiguation rule should apply. As the name suggests, the field *selected readings* will specify a proper subset of the readings that are specified in the field *readings filter*. The optional left and right context specifications constrain the environments under which the rule will apply. The effect of such a disambiguation rule can best be demonstrated by an example:[10]

(9) det, pron = det |adj*, noun|.

The rule in (9) applies to lexical tokens which have determiner and pronoun readings and retains only the determiner reading if the token is followed by any, including zero, number of adjectives and a noun.

While ODRs reduce the contextually valid readings for a single lexical node, DRRs simultaneously reduce readings of sequences of tokens. The latter type of rules is therefore used for simplifying the candidate morphological analyses of lexical nodes that make up phrasal categories.

The general format for DRRs is shown in (10).

(10) |node_sequence| ⇒ boolean_constraints.

(11) instantiates the DRR schema to the disambiguation rule needed for German to eliminate all readings of adjectives and nouns that do not match. The pattern matching algorithm of the XIP System will ensure non-deterministic application of the rule to each adjective that precedes a noun in a left-to-right fashion.

(11) |adj*, adj#1, adj*, noun#2 | ⇒ (#1[agr] :: #2[agr]).

The condition on the right-hand side of the rule (with the identity operator ::) enforces strict identity of agreement features between adjective and noun, with agreement consisting of the gender, number and case features for each node. Therefore, the rule has the effect of eliminating all readings of adjective and noun sequences with conflicting agreement features. However, if the nodes in question have no common readings to start with, then no readings are eliminated.

[9] If at a given level, more than one rule is applicable to a lexical node, the rule specified first in the rule file takes precedence.

[10] Compared to the actual rules used in the GRIP rule set, all ODRs and DRRs stated in this section have been simplified for expository purposes. The actual rules contain much more fine-grained context patterns that take into account what optional elements may intervene between the actual nodes that are being compared.

The rule in (12) accounts for the distinct declension class values required for contextually valid patterns of determiners and adjectives that we discussed in detail in section 2 above.

(12) |det#1, adj*, adj#2, adj*, noun| ⇒ (#1[agr] :: #2[agr]) & (#1[decl] ∼: #2[decl]).

If there is no determiner in front of a sequence of an adjective and a noun, then all weak readings of the adjective and the noun should be eliminated. This is handled by rule (13):

(13) |?[det:∼], adj*, adj#1, adj*, noun#2| ⇒ (#1[agr] :: #2[agr]) & (#1[decl: St]) & (#2[decl: St]).

Rules (12) and (13) illustrate another feature of the expressivity of DRRs in XIP: the constraint on the right-hand side of the DRR may contain any combination of Boolean operators (disjunction, conjunction and negation of features) that can be expressed in the system. To force distinctness of declension values the negated equality operator ∼: is used.

The full expressivity of DRRs makes it possible to state conditions on contextually valid morphological readings as succinctly as possible. This is one of the main advantages of the present approach over previous frameworks for morphological disambiguation.[11] While the framework of constraint grammar used by Voutilainen (1995) permits Boolean constraints, it lacks an equality operator and the use of variables over features on adjacent nodes. This, in turn, means that constraints cannot be generalized, but have to be stated in a case by case fashion. While this may be tolerable for languages like English, it will lead to an explosion of rules for languages like German with richer morphological paradigms.

Hajic (2001) and Oflazer (1997) do not consider agreement phenomena of the sort treated here. Therefore, it is difficult to tell whether the syntax of their disambiguation rules is rich enough to accommodate the same level of generality provided by the XIP DRRs.

Another important feature of XIP is that ODRs and DRRs can be freely mixed. In fact, mixing of the two rule types is often necessary. For example, as a result of an earlier application of DRRs, clauses often contain only one head noun that can be nominative. The other cases for this one noun can then be eliminated by an ODR. This reduction of readings on the head noun can, in turn, lead to a further reduction of the other lexical nodes (e.g. preceding determiners and adjectives) that belong to the same noun phrase.

5 Percolation of Morphological Features and Putting it all Together

As mentioned above, the dependency analysis takes as input the output of the chunk parser and tries to link nodes of the chunked tree by dependency relations. For example, chunked NPs are linked to the finite verb via grammatical relations such as *subject*, *direct object* and *indirect object*, depending on the morphological features present on the NP nodes. This section will explain how the contextually valid morphological analyses for the lexical nodes that make up an NP can be percolated up to the NP node during chunking.

The chunker uses non-recursive rewrite rules to combine the lexical nodes that make up an NP, after these nodes have been disambiguated by the use of ODRs and DRRs. The desired percolation of morphological features onto the mother node is carried out by side conditions on the rewrite rules that are specified by Boolean constraints analogous to those shown in the previous section for DRRs.

The resulting interaction between ODRs, DRRs and chunking rules can best be illustrated by the chunk analysis for one of the NPs of our original example (1). Fig. 4 shows the input to morphological disambiguation for the NP *verdienten Beamten*, with 25 candidate readings for the adjectives and 10 for the noun. The relevant DRR (13) eliminates non-shared readings and all weak readings for adjective and noun. Furthermore, the syntactic heuristic for conjoined NPs eliminates the non-dative readings for adjective and noun. As a result, the output of morphological disambiguation will retain only one analysis for

[11] Petkevic (2001) seems to envisage rules similar to the ones used in XIP. However, he does not provide any formal specification or semantics for disambiguation rules, which makes a precise comparison difficult.

each lexical node: `Adj+Masc+Pl+Dat+St` and `Noun+Masc+Pl+Dat+St`. The chunker then combines adjective and noun into an NP and percolates the agreement features of the remaining contextually valid reading onto this NP node.

Since XIP allows the inclusion of features on non-terminal nodes for chunking rules, readers might wonder why narrowing down contextually valid readings has to be done prior to chunking by the special-purpose mechanism of DRRs and could not, instead, be done during chunking by appropriate Boolean constraints on chunking rules. However, the latter is beyond the functionality of chunking rules, which do not allow to eliminate readings when forming chunks on the basis of feature values. Notice also that most chunk parsers do not allow the introduction of features on non-terminals and require, instead, that all non-terminals are atomic symbols. For such chunk parsers, the only option would be to create distinct non-terminal symbols for each combination of agreement values, resulting in a proliferation of phrasal and lexical categories and accompanying rule sets. The functionality of DRRs for such chunk parsers would therefore be at least as desirable as for XIP in order to reduce the processing load of chunk parsing.

6 Quantitative Evaluation

At present, GRIP contains a total of 106 DRRs which aim at morphological disambiguation of German noun phrases. Coverage of the rules includes prenominal agreement (with determiners, adjectives, cardinals, measure phrases, participial premodifiers, etc.), head-pronoun agreement for relative clauses, case agreement with prepositions, subject-verb agreement, agreement in complex proper names and titles, as well as simple nominal coordinations. The main reason for concentrating on noun phrase disambiguation is that it is the most crucial source of ambiguity for the subsequent assignment of dependency structure.

GRIP's morphological disambiguation component was evaluated on a corpus of 5732 tokens extracted from the `taz` newspaper corpus (taz, 1999). This test corpus contains a total of 1571 noun phrases. The corpus was automatically annotated by the XRCE morphological analyzer for German and then manually corrected so as to provide a gold standard for the present evaluation. The corpus has an average number of 3.68 distinct readings per token. 45.18% of all tokens are morphologically unambiguous. For lexical nodes that are contained in noun phrases, the average number of distinct readings is significantly higher: 6.08 per token; and only 13.00% of the nodes have a unique analysis. The fact that NPs exhibit a much higher than average degree of ambiguity further attests to the priority that has to be given to morphological disambiguation of NPs.

6.1 Morphological Disambiguation

DRRs apply to noun phrases with two or more lexical nodes. For this class of noun phrases, application of all DRRs results in an average of 1.55 contextually valid readings for the nominal head of the NP (compared to an average of 5.51 readings in the input). This corresponds to a 71.87% reduction of readings. Fig. 1 shows the distribution of the number of disambiguated readings for noun phrases with two or more lexical nodes.

	percentage
1 reading	58.65%
2 readings	34.31%
$\geq$ 3 readings	7.04%

Figure 1: Results of DRR Application

Thus, in 92.96% of all cases, at most two readings are retained, with more than half of all noun phrases uniquely disambiguated. However, for reliable assignment of dependency relations, a remaining ambiguity rate of more than 40% is not acceptable. Thus, further morphological disambiguation is necessary.

6.2 Adding Syntactic Heuristics

For the NPs that retain more than one valid analysis after DRR application, the syntactic environment in which they occur in the corpus can help to further disambiguate them. Notice also that DRRs will only apply to noun phrases consisting of more than one lexical node that exhibits inflectional morphology. DRRs will therefore not apply to single-element NPs such as relative or personal pronouns. In order to disambiguate such single-element NPs and to further disambiguate complex NPs, GRIP employs syntactic heuristics stated in the form of ODRs.

Description of Syntactic Heuristic	Case value	Percentage
The NP is the only one in a finite clause (then it is the single candidate for subject).	Nom	16.57%
A noun with feature `City` or `Country` is preceded by a preposition *in*.	Dat	4.07%
Eliminate Nom reading on ambiguous NPs if there is a non-ambiguous Nom NP in a clause (with no coordination or comparison).	¬ Nom	3.66%
The NP is an argument of a copula verb.	Nom	3.26%
A nominative reading does not agree with a finite verb in number.	¬ Nom	2.16%
The NP is neither preceded by a preposition nor by another NP.	¬ Gen	1.62%
The NP is a second (third) NP in a Vorfeld position in V2 clause.	Gen	1.21%
The NP is a complement of a `zu`-infinitive.	¬ Nom	1.09%

Figure 2: Syntactic Heuristics

One of the most effective syntactic heuristics that GRIP employs is to retain only the nominative case reading for an NP if that NP is the only candidate for being the subject (i.e. it is the only NP in a finite clause or the only NP with a nominative reading). In general, the form and contents of these rules is quite heterogeneous, and due to their heuristic nature, the rules may overapply in some cases. Manual inspection of the GRIP output of the test corpus revealed a total of 13 mistakes where lexical nodes contained in an NP did not retain the correct analysis. In all cases, these mistakes were due to the application of heuristic rules.

Fig. 2 provides an overview of some of the more effective heuristics currently implemented in GRIP. For each heuristic, Fig. 2 shows which case value is retained or eliminated. The numbers in Fig. 2 indicate the approximate percentage of ambiguous NPs that received a unique reading after the application of the heuristic.

	count of NPs	percentage
1 reading	1211	77.08%
2 readings	226	14.39%
≥ 3 readings	134	8.53%

Figure 3: Disambiguation after Application of DRRs and of Syntactic Heuristics for all NPs

Fig. 3 summarizes the results after application of all DRRs and of the full set of syntactic heuristics to all NPs.[12] Fig. 3 shows that in three out of four noun phrases, a unique reading can serve as input to the dependency parsing module of GRIP.[13]

Fig. 3 shows the distribution rates for all NPs. If one considers only NPs that contain more than one lexical node, then the disambiguation rate is even higher, as shown in Fig. 4. For this class of NPs more than eight out of ten NPs are uniquely disambiguated, and less than one percent retain more than two readings.

	percentage
1 reading	82.33%
2 readings	17.18%
≥ 3 readings	0.49%

Figure 4: Disambiguation after Application of DRRs and of Syntactic Heuristics for non-single-element NPs

Single-element NPs such as pronouns and proper names exhibit less inflectional variation than other nominal elements. Therefore, they are inherently ambiguous. DRRs, which compare two lexical nodes, do not apply to them. Since DRRs yield a much higher reduction in ambiguity rate compared to syntactic heuristics, it should therefore come as no surprise that single-element NPs remain ambiguous to a much higher degree. What this seems to show is that morphological disambiguation is not sufficient for disambiguation of single-element NPs.

[12] Since gender ambiguities never play a role in the determination of dependency relations, we disregard gender ambiguities in determining what counts as a unique reading.

[13] Currently, GRIP makes no attempt to disambiguate lexical nodes that do not belong to NPs. However, as a side effect of the application of syntactic heuristics to NPs, other lexical nodes (e.g. prepositions and verbs) are disambiguated in at least some cases. At present, 85.83% of all lexical nodes receive a unique analysis.

At this point we can only speculate on what techniques can be used. One promising strategy, at least for German, is to take into account the well-known ordering constraints among pronouns in the so-called *Wackernagel position* at the left edge of the Mittelfeld (Lenerz, 1977).[14] Other factors to consider are the thematic structure of the main verb and robust methods for anaphora resolution. However, it should be clear that such techniques go well beyond the realm of morpho-syntax and therefore have to be left to future research.

7 Conclusion

Morphological disambiguation constitutes a crucial step in narrowing down the search space for the correct assignment of dependency structures. A quantitative evaluation on a German test corpus has shown that application of XIP disambiguation rules yields unique morphological analyses as input for assigning dependency relations in 77.08% of all cases. For those NPs that still have multiple readings, the lexical resources (CELEX and IMS-LEX), which are used by the dependency parsing module of GRIP and which give subcategorization information for German verbs, can provide further constraints for disambiguation. A quantitative analysis of such disambiguation at the level of dependency parsing itself will be the subject of future research.

Acknowledgments

This research was conducted as part of the Sonderforschungsbereich 441 *Linguistic Data Structures* funded by the Deutsche Forschungsgemeinschaft. The authors deeply appreciate the contribution of Dr. Claude Roux of XRCE, Grenoble. He is the chief implementor of XIP and customized the disambiguation tools necessary for German morphological disambiguation. We also wish to thank Martina Liepert, Detmar Meurers, Kiril Simov, Tylman Ule, and Jorn Veenstra for their comments on earlier versions of this paper.

[14] For example, if two pronouns occur in the Wackernagel position, as in *dass sie sie sieht.* ('that she sees her'), then the nominative pronoun has to precede the accusative pronoun. Thus, even though both occurrences of *sie* are ambiguous between nominative and accusative case, the syntactic context can disambiguate their case specifications.

References

S. Ait-Mokhtar, J.-P. Chanod, and C. Roux. to appear. Robustness beyond Shallowness: Incremental Deep Parsing. *Natural Language Engineering*, Vol. 24.

D. Duchier. 1999. Axiomatizing Dependency Parsing Using Set Constraints. In: *Sixth Meeting on the Mathematics of Language (MOL6).* Orlando, Florida, pp. 115–126.

J. Hajic, P. Krbec, P. Květoň, K. Oliva, V. Petkevič. 2001. Serial Combination of Rules and Statistics: A Case Study in Czech Tagging. In: *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001)*, pp. 260–267.

E. Hinrichs, J. Bartels, Y. Kawata, V. Kordoni, and H. Telljohann. 2000. The Verbmobil Treebanks. In: *5. Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2000).* Ilmenau, pp. 107–112.

J. Lenerz 1977. *Zur Abfolge nominaler Satzglieder im Deutschen.* Max Niemeyer Verlag, Tübingen.

K. Oflazer and G. Tür. 1997. Morphological Disambiguation by Voting Constraints. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL 1997)*, pp. 122–129.

V. Petkevič. 2001. Grammatical Agreement and Automatic Morphological Disambiguation of Inflectional Languages. In: V. Matousek et al. (Eds.): In: *Proceedings of the International Conference on Text Speech and Dialogue (TSD 2001).* Lecture Notes in Artificial Intelligence, Vol. 2166. Springer Verlag, Berlin, pp. 47–53.

P. Tapanainen and T. JÆrvinen. 1997. A non-projective dependency parser. In: *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP'97).* ACL, Washington, D.C., pp. 64–71.

taz – die Tageszeitung (CD-ROM). 1999. September 1986 – May 1999. www.taz.de.

A. Voutilainen. 1995. Morphological Disambiguation. In: F. Karlsson et al. (Eds.) *Constraint Grammar.* Mouton de Gruyter, Berlin, pp. 165–285.

A. Zwicky. 1986. German adjective agreement in GPSG. *Linguistics*, Vol. 24, pp. 957–990.