When word order and part-of-speech tags are not enough — Swedish dependency parsing with rich linguistic features

Lilja Øvrelid
NLP-unit, Dept. of Swedish
Göteborg University
Sweden
lilja.ovrelid@svenska.gu.se

Joakim Nivre
Växjö University and
Uppsala University
Sweden
nivre@msi.vxu.se

Abstract

Even with high overall parsing accuracy, data-driven parsers often make errors in the assignment of core grammatical functions such as subject and object. Starting from a detailed error analysis of a state-of-the-art dependency parser for Swedish, we show that the addition of linguistically motivated features targeting specific error types may lead to substantial improvements, both for specific grammatical functions and in terms of overall parsing accuracy. In this way, we achieve the best reported results for dependency parsing of Swedish.

Keywords

data-driven parsing, dependency parsing, error analysis, grammatical relations, linguistic features, treebanks.

1 Introduction

Despite the dramatic improvement in accuracy for data-driven parsers in recent years, we still have relatively little knowledge about the exact influence of data-derived features on the parsing accuracy for specific linguistic constructions. A deeper analysis of specific error sources in data-driven parsing may therefore be one of the most important steps towards a further advancement of the state of the art.

There are a number of studies that investigate the influence of different features or representational choices on overall parsing accuracy, within a variety of different frameworks, e.g., [3], [12], [10], [2] and [8]. There are also attempts at a more fine-grained analysis of accuracy, targeting specific linguistic constructions or grammatical functions, such as [4], [6], and [11]. But there are few studies that combine the two perspectives and try to tease apart the influence of different features on the analysis of specific constructions, let alone motivated by a thorough linguistic analysis.

In this paper, we present an in-depth study of the influence of certain grammatical features, such as animacy, definiteness, and finiteness, on the parsing accuracy for core grammatical functions, in particular subjects and objects. The language analyzed is Swedish, which poses special problems for the identification of subjects and objects due to limited case marking and ambiguous word order patterns. The parsing framework is deterministic classifier-based dependency parsing, more precisely the MaltParser system [13], which achieved the highest parsing accuracy for Swedish in the CoNLL-X shared task on dependency parsing [5].

2 Parsing Swedish

Before we turn to a description of the treebank and the parser used in the experiments, we want to point to a few grammatical properties of Swedish that will be important in the following:

Verb second (V2) The finite verb always resides in second position in declarative main clauses.

Word order variation Pretty much any constituent may occupy the sentence-initial position. However, subjects are most common.

Limited case marking Nouns are only inflected for genitive case. Personal pronouns distinguish nominative and accusative case, but demonstratives and quantifying pronouns are case ambiguous (like nouns).

2.1 Treebank: Talbanken05

Talbanken05 is a Swedish treebank converted to dependency format, containing both written and spoken language [14].[1] For each token, Talbanken05 contains information on word form, part of speech, head and dependency relation, as well as various morphosyntactic and/or lexical semantic features. The nature of this additional information varies depending on part of speech:

NOUN:	*definiteness, animacy, case* (Ø/GEN)
PRO:	*pronoun type, animacy, case* (Ø/ACC)
ADJ:	grade of *comparison*
ADV:	*semantic class*, e.g., temporal
CONJ:	*semantic class*, e.g., disjunctive

2.2 Parser: MaltParser

We use the freely available MaltParser,[2] which is a language-independent system for data-driven dependency parsing. MaltParser is based on a deterministic parsing strategy, first proposed by Nivre (2003), in combination with treebank-induced classifiers for

[1] The written sections of the treebank consist of professional prose and student essays and amount to 197,123 running tokens, spread over 11,431 sentences.

[2] http://w3.msi.vxu.se/users/nivre/research/MaltParser.html

	FORM	POS	DEP	FEATS
S:*top*	+	+	+	+
S:*top*+1		+		
I:*next*	+	+		+
I:*next*−1	+			+
I:*next*+1	+	+		+
I:*next*+2		+		
G: head of *top*	+			+
G: left dep of *top*			+	
G: right dep of *top*			+	
G: left dep of *next*	+		+	+
G: left dep of head of *top*			+	
G: left sibling of right dep of *top*			+	
G: right sibling of left dep of *top*	+			+
G: right sibling of left dep of *next*		+	+	

Table 1: *Baseline and extended (*FEATS*) feature model for Swedish; S: stack, I: input, G: graph;* $\pm n$ *=* n *positions to the left(−) or right (+)*

predicting the next parsing action. Classifiers can be trained using any machine learning approach, but the best results have so far been obtained with support vector machines, using LIBSVM [7]. MaltParser has a wide range of parameters that need to be optimized when parsing a new language. As our baseline, we use the settings optimized for Swedish in the CoNLL-X shared task [15], and the only parameter that will be varied in the later experiments is the feature model used for the prediction of the next parsing action. Hence, we need to describe the feature model in a little more detail.

MaltParser uses two main data structures, a stack (S) and an input queue (I), and builds a dependency graph (G) incrementally in a single left-to-right pass over the input. The decision that needs to be made at any point during this derivation is (a) whether to add a dependency arc (with some label) between the token on top of the stack (*top*) and the next token in the input queue (*next*), and (b) whether to pop *top* from the stack or push *next* onto the stack. The features fed to the classifier for making these decisions naturally focus on attributes of *top*, *next* and neighbouring tokens in S, I or G. In the baseline feature model, these attributes are limited to the word form (FORM), part of speech (POS), and dependency relation (DEP) of a given token, but in later experiments we will add other linguistic features (FEATS). The baseline feature model is depicted as a matrix in Table 1, where rows denote tokens in the parser configuration (defined relative to S, I and G) and columns denote attributes. Each cell containing a + corresponds to a feature of the model.

3 Baseline and Error Analysis

The written part of Talbanken05 was parsed employing the baseline feature model detailed above, using 10-fold cross validation for training and testing. The overall result for unlabeled and labeled dependency accuracy is 89.87 and 84.92 respectively.[3]

Error analysis shows that the overall most frequent errors in terms of dependency relations involve either various adverbial relations (due to PP-attachment ambiguities and a large number of adverbial labels) or the core argument relations of subject and direct object. In particular, confusion of the two argument functions are among the top ten most frequent error types with respect to dependency assignment. The first three columns of Table 5 show confusion matrices for the assignment of the subject and direct object dependency relations.

The sources of errors in subject/object assignment are various. Common to all of these is that the parts of speech that realize subjects and objects are compatible with a range of dependency relations. Swedish, however, in addition exhibits ambiguities in morphology and word order which complicate the picture further. We will exemplify these factors through an analysis of the errors where subjects are assigned object status (SS_OO) and vice versa (OO_SS).

The confusion of subjects and objects follows from lack of sufficient formal disambiguation, i.e., simple clues such as word order, part-of-speech and word form do not clearly indicate syntactic function. The reason for this can be found in ambiguities on several levels.

With respect to word order, subjects and objects may both precede or follow their verbal head. Subjects, however, are more likely to occur preverbally (77%), whereas objects typically occupy a postverbal position (94%). Based on word order alone we would expect postverbal subjects and preverbal objects to be more dominant among the errors than in the treebank as a whole (23% and 6% respectively), since they display word order variants that depart from the canonical ordering of arguments. Table 2 shows a breakdown of the errors for confused subjects and objects and their position with respect to the verbal head. We find that postverbal subjects (after) are in clear majority among the subjects erroneously assigned the object relation. Due to the V2 property of Swedish, the subject must reside in the position directly following the finite verb whenever another constituent occupies the preverbal position, as in (1) where a direct object resides sentence-initially:

Gold	Sys	before	after	Total
SS	OO	103 (23.1%)	343 (76.9%)	446 (100%)
OO	SS	103 (33.3%)	206 (66.7%)	309 (100%)

Table 2: *Position relative to verb for confused subjects and objects*

(1) Samma erfarenhet gjorde **engelsmännen**
same experience made englishmen-DEF
'The same experience, the Englishmen had'

For the confused objects we find a larger proportion of preverbal elements than for subjects, which is the mirror image of the normal distribution of syntactic functions among preverbal elements. As Table 2 shows, the proportion of preverbal elements among the subject-assigned objects (33.3%) is notably higher than in the corpus as a whole, where preverbal objects account for a miniscule 6% of all objects.

In addition to the word order variation discussed above, Swedish also has limited morphological marking of syntactic function. Nouns are marked only for genitive case and only pronouns are marked for accusative case. There is also some syncretism in the pronominal paradigm where the pronoun is invariant

[3] Note that these results are slightly better than the official CoNLL-X shared task scores (89.50/84.58), which were obtained using a single training-test split, not cross-validation. Note also that, in both cases, the parser input contained gold standard part-of-speech tags.

Gold	Sys	Noun		Pro_{amb}		Pro_{unamb}		Other		Total	
SS	OO	324	72.6%	53	11.9%	29	6.5%	40	9.0%	446	100%
OO	SS	215	69.6%	74	23.9%	9	2.9%	11	3.6%	309	100%

Table 3: *Parts of speech for confused subjects and objects*

for case, e.g. *det, den* 'it', *ingen/inga* 'no', and may, in fact, also function as a determiner. This means that, with respect to word form, only the set of unambiguous pronouns clearly indicate syntactic function. We may predict that subject/object confusion errors frequently exhibit elements whose syntactic category and/or lexical form does not disambiguate, i.e., nouns or ambiguous pronouns. Table 3 shows the distribution of nouns, functionally ambiguous and unambiguous pronouns and other parts of speech for confused subjects/objects. Indeed, we find that nouns and functionally ambiguous pronouns dominate the errors where subjects and objects are confused.

The initial error analysis shows that the confusion of subjects and objects constitutes a frequent and consistent error during parsing. It is caused by ambiguities in word order and morphological marking and we find cases that deviate from the most frequent word order patterns and are not formally disambiguated by part-of-speech information. It seems clear that we in order to resolve these ambiguities have to examine features beyond syntactic category and linear word order.

4 Grammatical Features for Argument Disambiguation

The core arguments themselves tend to differ along several dimensions. The property of *animacy*, a referential property of nominal elements, has been argued to play a role in argument realization in a range of languages [1], [9]. It is closely correlated with the semantic property of agentivity, hence subjects will tend to be referentially animate more often than objects. Another property which may differentiate between the argument functions of subject and object is the property of *definiteness*, which can be linked with a notion of givenness [1], [17]. This is reflected in the choice of referring expression for the various argument types in Talbanken05 – subjects are more often pronominal (49.2%), whereas objects are typically realized by an indefinite noun (67.6%). The error analysis made clear the importance of not only distinguishing between the core arguments but also between arguments and non-arguments, and in particular determiners. Both the set of case ambiguous pronouns and a group of common nouns may function as determiners. The grammatical dimensions of *person* (1st/2nd vs 3rd), as well as *case* marking for nouns (genitive) are properties which may be beneficial in this respect.

As mentioned in section 2, there are categorical constraints which are characteristic for Swedish word order. Only subjects may follow a finite verb and precede a non-finite verb and only objects may occur after a non-finite verb. Information on *finiteness* is therefore something that one might assume to be beneficial for subject/object assignment. Another property of the verb which clearly influences the assignment of core argument functions is the *voice* of the verb, i.e., whether it is passive or active.

	Unlabeled	Labeled
Baseline	89.87	84.92
Pers	89.93	85.10
Def	89.87	85.02
Pro	89.91	85.04
Case	89.99	85.13
Verb	90.15	85.28
Pers&Def&Pro&Case	90.17	85.45
Pers&Def&Pro&Case&Verb	90.42	85.73
All	90.73	86.32

Table 4: *Overall results expressed as average unlabeled and labeled attachment scores*

5 Experiments

In the following we will experiment with the addition of morphosyntactic and lexical semantic features that approximate the distinguishing properties of the core argument functions discussed in section 4. We will isolate features of the arguments and the verbal head, as well as combinations of these, and evaluate their effect on overall parsing results as well as on subject/object disambiguation specifically.

5.1 Experimental methodology

All parsing experiments are performed using 10-fold cross-validation for training and testing on the entire written part of Talbanken05. The feature model used throughout is the extended feature model depicted in Table 1, including all four columns.[4] Hence, what is varied in the experiments is only the information contained in the FEATS features (animacy, definiteness, etc.), while the tokens for which these features are defined remains constant.

Overall parsing accuracy will be reported using the standard metrics of *labeled attachment score* (LAS) and *unlabeled attachment score* (UAS).[5] Statistical significance is checked using Dan Bikel's randomized parsing evaluation comparator.[6]

Since the main focus of this article is on the disambiguation of grammatical functions, we report accuracy for specific dependency relations, measured as a balanced F-score. We also employ two different comparative measures to compare parsers with respect to specific error types: (i) the number of errors of a certain type for the compared parsers (cf. Table 5), and (ii) the intersection of the errors of a certain type for the compared parsers.

5.2 Individual features

Talbanken05 explicitly distinguishes between person- and non-person referring nominal elements, a distinc-

[4] Preliminary experiments showed that it was better to tie FEATS features to the same tokens as FORM features (rather than POS or DEP features). Backward selection from this model was tried for several different instantiations of FEATS but with no significant improvement.

[5] LAS and UAS report the percentage of tokens that are assigned the correct head *with* (labeled) or *without* (unlabeled) the correct dependency label, calculated using eval.pl with default settings (http://nextens.uvt.nl/~conll/software.html)

[6] http://www.cis.upenn.edu/~dbikel/software.html

Confusion matrix for subjects (SS)

sys	Baseline #	Baseline % of tot.	Pers # (%)	Def # (%)	Pro # (%)	Case # (%)	Verb # (%)	PDPC # (%)	PDPCV # (%)	All # (%)
OO	446	25.9	388(13.0)	425(4.7)	401(10.1)	419(6.1)	365(18.2)	361(19.1)	293(34.3)	296(33.6)
ROOT	265	15.4	270(-1.9)	284(-7.2)	275(-3.8)	277(-4.5)	260(1.9)	269(-1.5)	266(-0.4)	241(9.1)
DT	238	13.8	196(17.6)	230(3.4)	218(8.4)	205(13.9)	239(-0.4)	164(31.1)	160(32.8)	160(32.8)
SP	206	12.0	203(1.5)	187(9.2)	198(3.9)	201(2.4)	216(-4.9)	188(8.7)	187(9.2)	195(5.3)
CC	137	8.0	135(1.5)	123(10.2)	139(-1.5)	139(-1.5)	122(10.9)	120(12.4)	114(16.8)	98(28.5)
FS	133	7.7	141(-6.0)	148(-11.3)	148(-11.3)	154(-15.8)	151(-13.5)	147(-10.5)	153(-15.0)	155(-16.5)
PA	53	3.1	53(0.0)	43(18.9)	43(18.9)	37(30.2)	49(7.5)	25(52.8)	22(58.5)	26(50.9)
...	...		...	...	...	...	...	...	...	

Confusion matrix for objects (OO)

sys	Baseline #	Baseline % of tot.	Pers # (%)	Def # (%)	Pro # (%)	Case # (%)	Verb # (%)	PDPC # (%)	PDPCV # (%)	All # (%)
SS	309	23.8	263(14.9)	288(6.8)	280(9.4)	273(11.7)	259(16.2)	251(18.8)	215(30.4)	212(31.4)
ROOT	221	17.0	239(-8.1)	224(-1.4)	237(-7.2)	229(-3.6)	218(1.4)	251(-13.6)	245(-10.9)	241(-9.0)
PA	126	9.7	122(3.2)	129(-2.4)	123(2.4)	112(11.1)	123(2.4)	111(11.9)	109(13.5)	105(16.7)
AA	103	7.9	94(8.7)	97(5.8)	92(10.7)	106(-2.9)	102(1.0)	96(6.8)	95(7.8)	74(28.2)
DT	99	7.6	95(4.0)	94(5.1)	99(0.0)	85(14.1)	99(0.0)	81(18.2)	70(29.3)	72(27.3)
ET	58	4.5	54(6.9)	61(-5.2)	57(1.7)	59(-1.7)	64(-10.3)	49(15.5)	49(15.5)	49(15.5)
OA	57	4.4	59(-3.5)	58(-1.8)	58(-1.8)	57(0.0)	65(-14.0)	63(-10.5)	66(-15.8)	64(-12.3)
...	...		...	...	...	...	...	...	...	

Table 5: *Confusion matrices for the assignment of the subject and object dependency relations for the baseline parser (columns 2–3) and for the different extended feature models (columns 4–11). For the baseline parser, we give the absolute number of occurrences of each error type, together with the percentage of each error type out of all subject/object errors. For the extended parsers, we give absolute numbers (#) along with relative improvement compared to the baseline (%)*

tion which overlaps fairly well with the traditional notion of **animacy**. As Table 4 shows, the addition of information on animacy for nominal elements causes an improvement in overall results (p<.0002). The subject and object functions are the dependency relations whose assignment improves the most when animacy information is added. There is also an effect for a range of other functions where animacy is not directly relevant, but where the improved analysis of arguments contributes towards correct identification (e.g., adverbials and determiners). If we take a closer look at the individual error types involving subjects and objects in Table 5, we find that the addition causes a reduction of errors confusing subjects with objects (SS_OO), determiners (SS_DT) and subject predicatives (SS_SP) – all functions which do not embody the same preference for person reference as subjects. The intersection of errors confusing subjects and objects shows that we improve on 23.1% of the SS_OO errors and 28.8% of the OO_SS errors made by the baseline parser when adding information on animacy.

Morphological **definiteness** is marked for all common nouns in Talbanken05. The addition of information on definiteness during parsing causes a slight (at the p<.03 level) improvement of overall results. Most noteworthy is an improvement in the identification of subject predicatives (SP), which are often confused with subjects (cf. Table 5). Nominal predicatives in Swedish are often realized by an indefinite noun (89.4%).

The addition of information on **pronoun type**[7] causes a general improvement in overall parsing results (p<.01), as we can see from Table 4. The dependency relations whose assignment improves the most are, once again, the core argument functions (SS, OO), as well as determiners (DT). We also find a general improvement in terms of recall for the assignment of the formal subject (FS) and object (FO) functions, which are both realized by the third person neuter pronoun *det* 'it', annotated as impersonal in the treebank.

Talbanken05 contains morphological **case** annotation for pronouns (null/accusative) and common nouns (null/genitive). Whereas we noted in the initial error analysis that case marking is not sufficient to disambiguate the targeted errors, we observed that core arguments were confused for determiners due to ambiguity in syntactic category and word form. When we employ case information during parsing we find a clear improvement in results (p<.0001). However, the improvement is not first and foremost caused by improvement in assignment of subjects and objects, but rather, the assignment of determiners and prepositional objects.

Talbanken05 contains morphosyntactic information on **tense** and **voice** for all verbs. In this experiment, all information available for the verbal category is included during parsing. As Table 4 shows, the addition of morphosyntactic information for verbs causes a clear improvement in overall results (p<.0001). The added information has a positive effect on the verbal dependency relations for finite and non-finite verbs, as well as an overall effect on the assignment of subjects and objects. Information on voice also benefits the relation expressing the demoted agent (AG) in passive constructions. The overview of individual error types typically involved in the assignment of the core argument functions (cf. confusion matrices in Table 5) indicates that the addition of information on verbal features improves on the confusion of the main argument types – subjects and objects (SS_OO, OO_SS), as well as subjects and expletive subjects (SS_FS). With respect to the intersection of errors performed by the two parsers confusing subjects and objects, we observe an improvement of 33.2% (SS_OO) and 37.2% (OO_SS) for the parser with added verbal features.

[7] There are 12 pronoun types in Talbanken05 which differentiate between, e.g., local (1st/2nd) and 3rd person pronouns, reflexive, reciprocal, interrogative, impersonal pronouns, etc.

Deprel		Freq	Baseline	PDPCV
SS	subject	0.1105	91.37	92.73
OO	object	0.0632	85.83	87.62
DT	determiner	0.1081	95.49	96.42
SP	subject predicative	0.0297	85.47	86.72
FS	formal subject	0.0050	71.81	74.57
PA	prep argument.	0.1043	95.03	95.74

Table 6: *Comparison of balanced F-scores for the core argument relations in the combined experiment (PDPCV).*

5.3 Feature combinations

The following experiments combine different nominal argument features, nominal argument features with verbal features, and finally all available grammatical features in Talbanken05.

The combination of the argument features of animacy, definiteness, pronoun type and case (PDPC), as well as the addition of verbal features to this feature combination (PDPCV) causes a clear improvement compared to the baseline *and* each of the individual feature experiments (p<.0001) (cf. Table 4). Since the results are better than the individual runs, we may conclude that there is a cumulative effect of the combined information.

Table 6 shows a comparison of the balanced F-scores for the argument dependency relations in the baseline and PDPCV experiments. If we examine the counts for individual error types in Table 5, we find an error reduction for the confused subjects with objects and vice versa with 34.3% and 30.4% respectively. With respect to the specific errors performed by the baseline parser for this error type, and targeted by the experiments, we observe a reduction of 45.7% for SS_OO and 46.6% for OO_SS.

When we add the remaining grammatical features in Talbanken05, i.e., the features for adjectives, adverbs, conjunctions and subjunctions, we observe an improvement (p<.0001) for the conjunct relation as well as the argument functions (SS, OO), determiners, verbal relations and adverbials. If we examine the intersected errors performed by the baseline parser in terms of confused subjects and objects, we find an improvement of 53.4% for the SS_OO error type and 50.5% for the OO_SS.

6 Conclusion

An in-depth error analysis of the best performing data-driven dependency parser for Swedish revealed consistent errors in dependency assignment, namely the confusion of core argument functions, resulting from word order ambiguity and lack of case marking. A set of experiments were designed to examine the effect of various linguistically motivated grammatical features hypothesized to target these errors.

The experiments showed that each feature individually caused an improvement in terms of overall labeled accuracy, performance for the core argument relations, and error reduction for the specific types of errors performed by the baseline parser. In particular, the final experiment (All), exhibited an error reduction of about 50% for the errors specifically targeted following the initial error analysis. In this way, we have also advanced the state of the art in Swedish dependency parsing, increasing the labeled accuracy of the best performing parser by 1.4 percentage points.

A possible objection to the applicability of the results presented above is that the added information consists of gold standard annotation from a treebank. However, the morphosyntactic features examined here are for the most part straightforwardly derived (definiteness, case, person, tense, voice) and represent standard output from most part-of-speech taggers. The property of animacy has been shown to be fairly robustly acquired for common nouns by means of distributional features from a shallow-parsed corpus [16].

Specific plans for future work relate to further error analysis of the baseline parser, including other non-argument relations, most notably adverbials, and similar experiments to more fully understand the interplay of the various features. On a more general note, the development of methods for in-depth error analysis which relate to specific linguistic constructions constitutes an important direction for gaining further knowledge about the types of generalizations acquired through data-driven syntactic parsing.

References

[1] J. Aissen. Differential Object Marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21:435–483, 2003.

[2] D. M. Bikel. Intricacies of Collins' parsing model. *Computational Linguistics*, 30(4):479–511, 2004.

[3] R. Bod. *Beyond Grammar*. CSLI Publications. University of Chicago Press, 1998.

[4] S. Buchholz. *Memory-Based Grammatical Relation Finding*. PhD thesis, Tilburg University, 2002.

[5] S. Buchholz and E. Marsi. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 149–164, 2006.

[6] J. Carroll and E. Briscoe. High precision extraction of grammatical relations. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pages 134–140, 2002.

[7] C.-C. Chang and C.-J. Lin. Libsvm: A library for Support Vector Machines. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm, 2001.

[8] E. Charniak and M. Johnson. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 173–180, 2005.

[9] Ö. Dahl and K. Fraurud. Animacy in grammar and discourse. In T. Fretheim and J. K. Gundel, editors, *Reference and referent accessibility*, pages 47–65. John Benjamins, Amsterdam, 1996.

[10] D. Klein and C. D. Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 423–430, 2003.

[11] S. Kübler and J. Prokić. Why is German dependency parsing more reliable than constituent parsing? In *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (TLT)*, pages 7–18, 2006.

[12] B. Megyesi. Shallow parsing with PoS taggers and linguistic features. *Journal of Machine Learning Research*, 2:639–668, 2002.

[13] J. Nivre, J. Hall, and J. Nilsson. MaltParser: A data-driven parser-generator for dependency parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, pages 2216–2219, 2006.

[14] J. Nivre, J. Nilsson, and J. Hall. Talbanken05: A Swedish treebank with phrase structure and dependency annotation. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC2006)*, Genoa, Italy, May 24-26 2006.

[15] J. Nivre, J. Nilsson, J. Hall, G. Eryiğit, and S. Marinov. Labeled pseudo-projective dependency parsing with Support Vector Machines. In *Proceedings of the tenth conference on Computational Natural Language Learning (CoNLL)*, 2006.

[16] L. Øvrelid. Towards robust animacy classification using morphosyntactic distributional features. In *Proceedings of the EACL 2006 Student Research Workshop*, Trento, Italy, 2006.

[17] A. Weber and K. Müller. Word order variation in German main clauses: A corpus analysis. In *Proceedings of the 20th International conference on Computational Linguistics*, pages 71–77, 2004.