Lithuanian Dependency Parsing with Rich Morphological Features

Jurgita Kapočiūtė-Dzikienė
Kaunas University of Technology
K. Donelaičio 73
LT-44249 Kaunas, Lithuania
jurgita.k.dz@gmail.com

Joakim Nivre
Uppsala University
Box 635
SE-75126 Uppsala, Sweden
joakim.nivre@lingfil.uu.se

Algis Krupavičius
Kaunas University of Technology
K. Donelaičio 73
LT-44249 Kaunas, Lithuania
pvai@ktu.lt

Abstract

We present the first statistical dependency parsing results for Lithuanian, a morphologically rich language in the Baltic branch of the Indo-European family. Using a greedy transition-based parser, we obtain a labeled attachment score of 74.7 with gold morphology and 68.1 with predicted morphology (77.8 and 72.8 unlabeled). We investigate the usefulness of different features and find that rich morphological features improve parsing accuracy significantly, by 7.5 percentage points with gold features and 5.6 points with predicted features. As expected, CASE is the single most important morphological feature, but virtually all available features bring some improvement, especially under the gold condition.

1 Introduction

During the last decades, we have seen a tremendous increase in the number of syntactic parsers available for different languages, often enabled by the development of syntactically annotated corpora, or treebanks. The added linguistic diversity has highlighted the fact that typological differences between languages lead to new challenges, both in parsing technology and treebank annotation. In particular, it has been observed repeatedly that richly inflected languages, which often also exhibit relatively free word order, usually obtain lower parsing accuracy, especially compared to English (Buchholz and Marsi, 2006; Nivre et al., 2007). This has led to a special interest in parsing methods for such languages (Tsarfaty et al., 2010; Tsarfaty et al., 2013).

In this paper, we contribute to the growing pool of empirical evidence by presenting the first statistical dependency parsing results for Lithuanian, a morphologically rich Baltic language characterized as one of the most archaic living Indo-European languages (Gimbutas, 1963).

Using the newly developed Lithuanian Treebank, we train and evaluate a greedy transition-based parser and in particular investigate the impact of rich morphological features on parsing accuracy. Our experiments show that virtually all morphological features can be beneficial when parsing Lithuanian, which contrasts with many previous studies that have mainly found a positive impact for isolated features such as CASE (Eryigit et al., 2008). Using all available features, we achieve a labeled attachment score of 74.7 with gold morphology (including part-of-speech tags and lemmas) and 68.1 with predicted morphology. The corresponding unlabeled attachment scores are 77.8 and 72.8, respectively.

2 The Lithuanian Treebank

The *Lithuanian Treebank* was developed by the Center of Computational Linguistics, Vytautas Magnus University.[1] The annotated texts are taken from the newspaper domain and thus represent normative Lithuanian language. The treebank contains 1,566 sentences and 24,265 tokens: 19,625 words (9,848 distinct) plus 4,640 punctuation marks (12 distinct). Word tokens in the Lithuanian Treebank are mor-

[1]The treebank creation was one of the tasks of the project *Internet Resources: Annotated Corpus of the Lithuanian Language and Tools of Annotation*, implemented in 2007-2008 and funded by the Lithuanian Science and Studies Foundation.

Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically Rich Languages, pages 12–21,
Seattle, Washington, USA, 18 October 2013. ©2013 Association for Computational Linguistics

	SBJ	OBJ	MODIF	PRED	ATTR	DEP	ROOT	TOTAL
Abbreviation	6					457	22	485
Acronym						31	2	33
Adjectival participle	1		28			84	12	125
Adjective	1			63	1,104	157	75	1,400
Adverbial participle			37			28	3	68
Adverb			1,134			193	29	1,356
Conjunction		5				1,171	93	1,269
Infinitive		6		372	9	139	21	547
Interjection						3	6	9
Noun	775	1,097	1,314		1,712	1,415	217	6,530
Numeral		1	22		158	72	6	259
Participle	1			150	430	285	197	1,063
Particle			27	78	1	216	36	358
Preposition		253	168			630	35	1,086
Pronoun	258	170	104		558	424	21	1,535
Proper noun	15	1	22		20	1,307	60	1,425
Roman number						25	3	28
Verb						205	1,844	2,049
TOTAL	1,057	1,533	2,856	663	3,992	6,842	2,682	19,625

Table 1: Cooccurrence statistics on dependencies (columns) and PoS tags (rows) in the Lithuanian Treebank.

phologically and syntactically annotated as follows:

- Syntactic dependencies: 7 different categories listed in Table 1 (columns).
- Part-of-Speech (PoS) tags: 18 different categories listed in Table 1 (rows). These tags simply determine PoS but do not incorporate any additional morphological information.
- Morphological features: 12 different categories listed with possible values in Table 2. The number of morphological features assigned to a word varies from 0 (for particles, conjunctions, etc.) to 9.[2]
- Lemmas: base form of word, lowercase except for proper names.

The syntactic annotation scheme only distinguishes 5 basic grammatical relations (SBJ, OBJ, PRED, ATTR, MODIF) plus an additional underspecified relation (DEP) for other dependencies between words and a special relation (ROOT) for words attached to an (implicit) artificial root node. The dependency structure always forms a tree originating from the root node, but there may be more than one token attached to the root node. This happens when a sentence contains several clauses which do not share any constituents. Table 1 gives statistics on the different dependency relations and their distribution over different PoS tags.

Examples of syntactically annotated sentences are presented in Figure 1 and Figure 2. All dependency relations are represented by arrows pointing from the head to the dependent, the labels above indicate the dependency type.[3] For example, as we can see in Figure 1, *nerizikuoja* (does not risk) is the head of *Kas* (Who) and this dependency relation has the SBJ label. The sentence in Figure 1 contains two clauses (separated by a comma) both containing SBJ dependency relations. The sentence in Figure 2 contains the main clause *Bet štai pro medį praslinko nedidelis šešėlis* and the subordinate clause *kuriame sėdėjau* in which the subject is expressed implicitly (a pronoun *aš* (I) can be inferred from the singular 1st person inflection of the verb *sėdėjau* (sat)). In Lithuanian sentences, the subject is very often omitted, and even the verb can be expressed implicitly. For exam-

[2] For example, the participle *esanti* (existent) is described by 8 feature values: CASE: Nominative, GENDER: Feminine, NUMBER: Singular, TENSE: Present, VOICE: Active, REFLEX: Non-reflexive, PRONOM: Non-pronominal, ASPECT: Positive.

[3] ROOT dependencies are not shown explicitly.

Category	Values	Frequency	Compatible PoS Tags
CASE	Nominative	3,421	Adjective, Noun, Numeral, Participle, Pronoun, Proper noun
	Genitive	4,204	
	Dative	445	
	Accusative	1,995	
	Instrumental	795	
	Locative	849	
	Vocative	10	
GENDER	Masculine	7,074	Adjective, Adverbial participle, Noun, Numeral, Participle, Pronoun, Proper noun
	Feminine	4,482	
	Neuter	283	
	Appellative	1	
NUMBER	Singular	8,822	Adjective, Adverbial participle, Noun, Numeral, Participle, Pronoun, Proper noun, Verb
	Plural	4,624	
	Dual	3	
TENSE	Present	1,307	Adjectival participle, Participle, Verb
	Past occasion	1,352	
	Past	311	
	Past iterative	31	
	Future	123	
MOOD	Indicative	1,950	Verb
	Subjunctive	87	
	Imperative	12	
PERSON	1^{st}	281	Verb
	2^{nd}	41	
	3^{rd}	1,727	
VOICE	Active	456	Participle
	Passive	594	
	Gerundive	13	
REFLEX	Reflexive	526	Adjectival participle, Adverbial participle, Infinitive, Noun, Participle, Verb
	Non-reflexive	3,486	
DEGREE	Positive	1,712	Adjective, Adverb, Numeral, Participle
	Comparative	1,712	
	Superior	1	
	Superlative	94	
TYPE	Cardinal	145	Numeral
	Ordinal	105	
	Multiple	9	
PRONOM	Pronominal	247	Adjective, Participle, Pronoun, Numeral
	Non-pronominal	3,056	
ASPECT	Positive	6,206	Adjectival participle, Adjective, Adverbial participle, Adverb, Infinitive, Noun, Participle, Particle, Preposition, Verb
	Negative	422	

Table 2: Morphological categories in the Lithuanian Treebank: possible values, frequencies and compatible PoS tags.

ple, in the sentence *Jis geras žmogus* (He is a good man), the copula verb *yra* (is) is omitted.

The possible values of different morphological categories are presented with descriptive statistics in Table 2. Given that word order in Lithuanian sentences is relatively free, morphological information is important to determine dependency relations. For example, an adjective modifying a noun has to agree in GENDER, NUMBER and CASE, as in *gražus miestas* (beautiful city), where both the adjective and the noun are in masculine singular nominative. Verbs agree with their subject in NUMBER and PERSON, as in *jūs važiuojate* (you are going) in second person plural. Finally, the CASE of a noun

Figure 1: Annotated sentence from the Lithuanian Treebank, consisting of two independent main clauses. Translation: *Who does not risk, that does not drink champagne but does not cry tearfully either.*

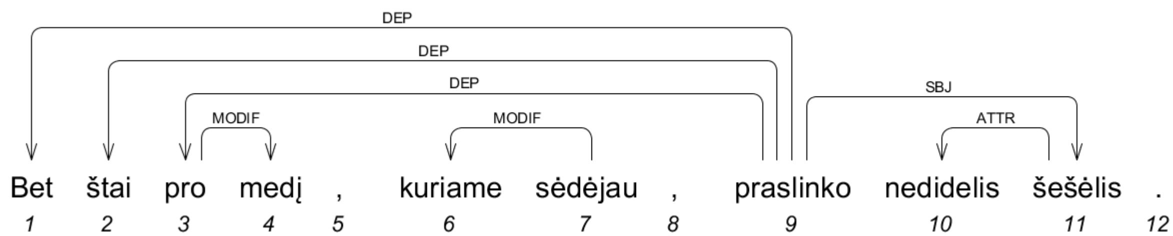

Figure 2: Annotated sentence from the Lithuanian Treebank, consisting of a main clause and a subordinate clause. Translation: *But here through the tree in which I sat passed a small shadow.*

or pronoun is an important indicator of the syntactic relation to the verb, such that nominative CASE almost always implies a SBJ relation. However, the transparency of morphological information is limited by syncretism in CASE, NUMBER and GENDER. Thus, the form *mamos* (mother(s)) can be either plural nominative or singular genitive; the form *mokytojas* (teacher(s)) can be either masculine singular nominative or feminine plural accusative.

3 Parsing Framework

We use the open-source system MaltParser (Nivre et al., 2006a) for our parsing experiments with the Lithuanian Treebank. MaltParser is a transition-based dependency parser that performs parsing as greedy search through a transition system, guided by a history-based classifier for predicting the next transition (Nivre, 2008). Although more accurate dependency parsers exist these days, MaltParser appeared suitable for our experiments for a number of reasons. First of all, greedy transition-based parsers have been shown to perform well with relatively small amounts of training data (Nivre et al., 2006b). Secondly, MaltParser implements a number of different transition systems and classifiers that can be explored and also supports user-defined input formats and feature specifications in a flexible way. Finally, MaltParser has already been applied to a wide range of languages, to which the results can be compared. In particular, MaltParser was used to obtain the only published dependency parsing results for Latvian, the language most closely related to Lithuanian (Pretkalniņa and Rituma, 2013).

In our experiments, we use the latest release of MaltParser (Version 1.7.2).[4] After preliminary experiments, we decided to use the arc-eager transition system (Nivre, 2003) with pseudo-projective parsing to recover non-projective dependencies (Nivre and Nilsson, 2005) and the LIBLINEAR learning package with multiclass SVMs (Fan et al., 2008). Table 3 lists the options that were explored in the preliminary experiments. We first tested all possible combinations of learning method and parsing algorithms and then performed a greedy sequential tuning of the options related to covered roots, pseudo-projective parsing, and all combinations of allow-root and allow-reduce.

In order to use MaltParser on the Lithuanian Treebank, we first converted the data to the CoNLL-X format,[5] treating all morphological feature bundles

[4] Available at http://maltparser.org.

[5] See http://ilk.uvt.nl/conll/#dataformat.

Option	Value
Learning method (-l)	liblinear
Parsing algorithm (-a)	nivreeager
Covered roots (-pcr)	head
Pseudo-projective parsing (-pp)	head+path
Allow root (-nr)	true
Allow reduce (-ne)	true

Table 3: List of MaltParser options explored in preliminary experiments with best values used in all subsequent experiments.

as a single string and putting it into the FEATS column, which means that there will be one boolean feature for each unique set of features. However, in order to study the influence of each individual morphological feature, we also prepared an appropriate format where every morphological feature had its own (atom-valued) column (called CASE, GENDER, NUMBER, etc.), which means that there will be one boolean feature for each unique feature value, as specified in Table 2. In the following, we will refer to these two versions as Set-FEATS and Atom-FEATS, respectively. Another choice we had to make was how to treat punctuation, which is not integrated into the dependency structure in the Lithuanian Treebank. To avoid creating spurious non-projective dependencies by attaching them to the root node, we simply attached all punctuation marks to an adjacent word.[6] Therefore, we also exclude punctuation in all evaluation scores.

We use five-fold cross-validation on the entire treebank in all our experiments. This means that the final accuracy estimates obtained after tuning features and other parameters may be overly optimistic (in the absence of a held-out test set), but given the very limited amount of data available this seemed like the most reasonable approach. We perform experiments under two conditions. In the Gold condition, the input to the parser contains PoS tags, lemmas and morphological features taken from the manually annotated treebank. In the Predicted condition, we instead use input annotations produced by the morphological analyser and lemmatizer *Lemuoklis* (Zinkevičius, 2000; Daudaravičius et al., 2007), which also solves morphological disambiguation problems at the sentence level. Table 4 shows the accuracy of this system for the output categories that are relevant both in the Set-FEATS and Atom-FEATS format.

[6]This is automatically handled by the covered roots option in MaltParser; see Table 3.

Category	Accuracy
POSTAG	88.1
LEMMA	91.1
Set-FEATS	78.6
Atom-FEATS	
CASE	87.2
GENDER	88.3
NUMBER	86.2
TENSE	94.1
MOOD	95.9
PERSON	95.8
VOICE	90.2
REFLEX	93.3
DEGREE	90.3
TYPE	80.7
PRONOM	89.3
ASPECT	93.5

Table 4: Accuracy of the morphological analyzer and lemmatizer used in the Predicted condition.

4 Parsing Experiments and Results

In our first set of experiments, we tuned two feature models in the Gold condition:

- **Baseline:** Starting from the default feature model in MaltParser, we used backward and forward feature selection to tune a feature model using only features over the FORM, LEMMA, POSTAG and DEPREL fields in the CoNLL-X format (that is, no morphological features). Only one feature was explored at a time, starting with FORM and going on to LEMMA, POSTAG, DEPREL, and conjunctions of POSTAG and DEPREL features. The best templates for each feature type were retained when moving on to the next feature.

- **Baseline+FEATS:** Starting from the Baseline model, we used forward feature selection to tune a feature model that additionally contains features over the FEATS field in the Set-FEATS

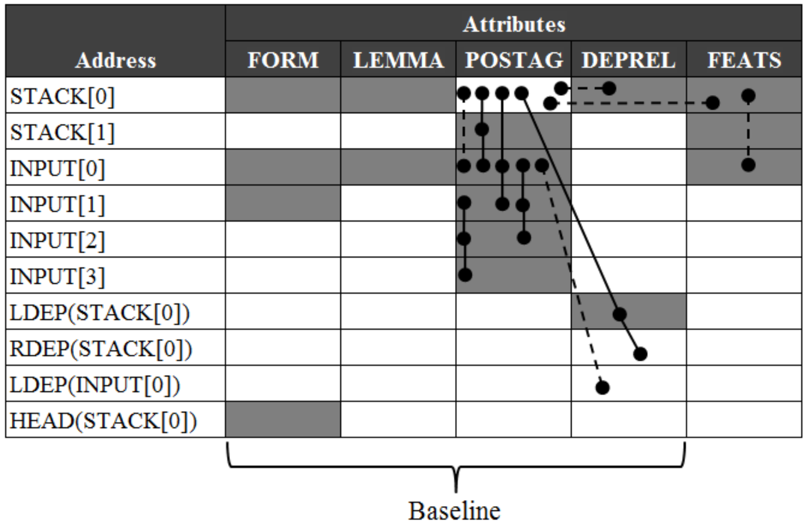

Figure 3: The feature models Baseline and Baseline+FEATS. Rows represent address functions, columns represent attribute functions. Gray cells represent single features, dotted lines connecting cell pairs or lines connecting cell triplets represent conjoined features. The Baseline model contains only features that do not involve the FEATS column.

version, optionally conjoined with POSTAG features.

The features included in these two models are depicted schematically in Figure 3. The Baseline+FEATS model includes all features, while the Baseline model includes all features except those that refer to the FEATS field. In the Gold condition, the Baseline model achieves a labeled attachment score (LAS) of 67.19 and an unlabeled attachment score (UAS) of 73.96, while Baseline+FEATS gets 74.20 LAS and 77.40 UAS. In the Predicted condition, the corresponding results are 62.47/70.30 for Baseline and 68.05/72.78 for Baseline+FEATS. Thus, with the addition of morphological features (all of them together) the Baseline+FEATS model exceeds the Baseline by 7.01 percentage points for LAS and 3.44 for UAS in the Gold condition and by 5.58 percentage points for LAS and 2.48 for UAS in the Predicted condition. To determine whether the differences are statistically significant we performed McNemar's test (McNemar, 1947) with one degree of freedom. The test showed the differences in LAS and UAS between Baseline and Baseline+FEATS for both the Gold and Predicted conditions to be statistically significant with $p << 0.05$.

In our second set of experiments, we started from the Baseline model and incrementally added morphological features in the Atom-FEATS format, one morphological category at a time, using the same five feature templates (three single and two conjoined) as for FEATS in the Baseline+FEATS model (see Figure 3). The order of explored morphological features was random, but only features that increased parsing accuracy when added were retained when adding the next morphological feature. The LAS results of these experiments are summarized in Figure 4 (reporting results in the Gold condition) and Figure 5 (in the Predicted condition). We do not present UAS results because they show the same trend as the LAS metric although shifted upwards. In the Gold condition, the best feature model is Baseline + CASE + GENDER + NUMBER + TENSE + DEGREE + VOICE + PERSON + TYPE, which achieves 74.66 LAS and 77.84 UAS and exceeds the Baseline by 7.47 percentage points for LAS and 3.88 for UAS (MOOD, REFLEX, PRONOM and ASPECT made no improvements or even degraded the performance). In the Predicted condition, the best feature model remains Baseline+FEATS, but using the Atom-FEATS version the best results are achieved with Baseline + CASE + GENDER + TENSE + VOICE + PERSON + REFLEX, which exceeds the Baseline by 5.36 percentage points for LAS and 2.55 for UAS (NUMBER, MOOD, DEGREE, REFLEX, PRONOM and ASPECT made no improvements or even degraded the performance). All these differences are statistically significant. By contrast, the differences between the best models with Atom-FEATS and Set-FEATS are not statistically significant for any metric or condition (with p values in the range 0.35–0.87).

5 Discussion

First of all, we may conclude that the Baseline feature model (without morphological information) does not perform very well for a morphologically rich language like Lithuanian (see Figure 4 and Figure 5), despite giving high accuracy for morphologically impoverished languages like English. However, it is likely that the accuracy of the Baseline model would be a bit higher for the Lithuanian Treebank if PoS tags incorporated some morphological information as they do, for example, in the English Penn Treebank (Marcus et al., 1993).

It thus seems that basic PoS tags as well as lemmas are too general to be beneficial enough for Lithuanian. The simple morphemic word form could be more useful (even despite the fact that Lithuanian is syncretic language), but the treebank is currently too small, making the data too sparse to create a robust model.[7] Thus, the effective way of dealing with unseen words is by incorporating morphological information.

In the Predicted condition, we always see a drop in accuracy compared to the Gold condition, although our case is not exceptional. For example, the Baseline model has a drop in LAS of 4.72 percentage points from Gold to Predicted, but this gap could possibly be narrowed by retuning the feature model for the Predicted condition instead of simply reusing the model tuned for the Gold condition. We also tried training the model on gold annotations for parsing predicted annotations, but these produced even worse results, confirming that it is better to make the training condition resemble the parsing condition. Despite noisy information, morphological features are still very beneficial compared to not using them at all (see Figure 5). Our findings thus agree with what has been found for Arabic by Marton et al. (2013) but seem to contradict the results obtained for Hebrew by Goldberg and Elhadad (2010).

As we can see from both curves in Figure 4 and Figure 5, the top contributors are CASE, VOICE, and TENSE, but the CASE feature gives the biggest contribution to accuracy. It boosts LAS by 6.51 points in the Gold condition and almost 5 points in the Predicted condition, whereas the contribution of all the other morphological features is less than 1 point (and not statistically significant). In a control experiment we reversed the order in which morphological features are added (presented in Figure 4 and Figure 5), adding CASE at the very end. In this case, the addition of all features except case resulted in a statistically significant improvement in the Gold condition ($p = 0.001$) but not in the Predicted condition ($p = 0.24$). However, the contribution of CASE was by far the most important again – increasing LAS by 5.55 points in the Gold condition and by 4.68 points in the Predicted condition. To further investigate the selection of morphological features, we also performed a greedy selection experiment. During this experiment CASE was selected first, again proving it to be the most influential feature. It was followed by VOICE, MOOD, NUMBER and DEGREE in the Gold condition and by GENDER, TENSE, PERSON and TYPE in the Predicted condition. Overall, however, greedy selection gave worse results than random selection, achieving 74.42 LAS and 77.60 UAS in the Gold condition and 67.83 LAS and 72.80 UAS in the Predicted condition.

To find that CASE is the most important feature is not surprising, as CASE has been shown to be the most helpful feature for many languages (at least in the Gold condition). But whereas few other features have been shown to help for other languages, in our case the majority of features (8 out of 12 in the Gold condition) are beneficial for Lithuanian. The so-called agreement features (GENDER, NUMBER and PERSON) are beneficial for Lithuanian (at least in the Gold condition) as well as for Arabic (Marton et al., 2013), but not such languages as Hindi (Ambati et al., 2010) and Hebrew (Goldberg and Elhadad, 2010). In the Predicted condition, their positive impact is marginal at best, possibly because NUMBER is very poorly predicted by the morpho-

[7] We tried to reduce data sparseness a little bit by changing all words into lowercase, but the drop in accuracy revealed that orthographic information is also important for parsing.

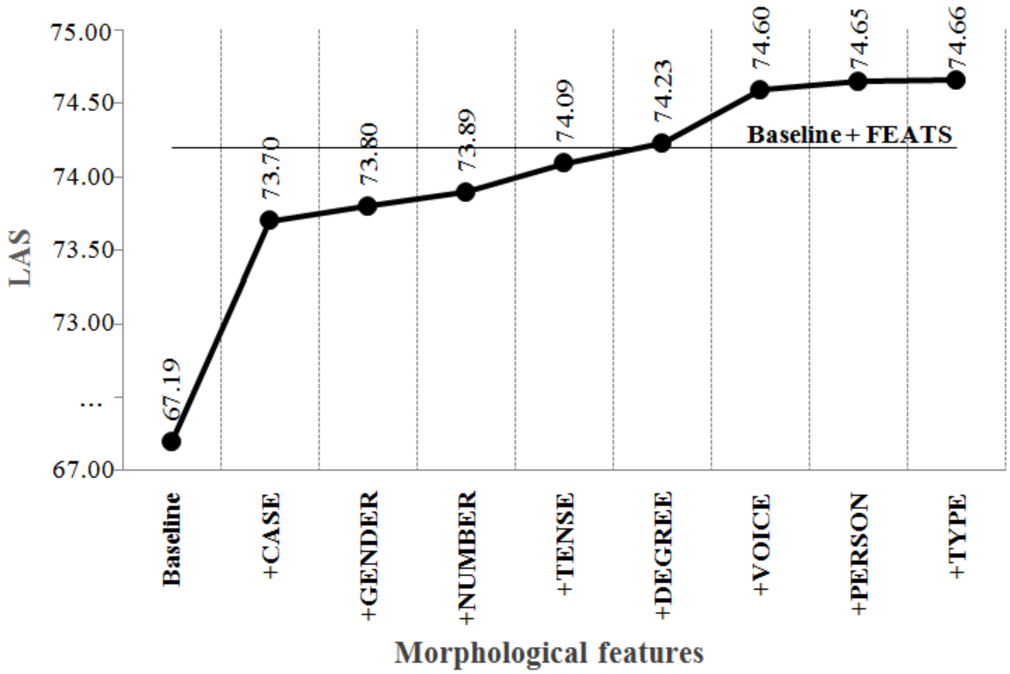

Figure 4: The contribution of individual morphological features in the Gold condition. The x axis represents feature models incorporating different attributes; the y axis represents LAS. The horizontal line at 74.20 represents the LAS of Baseline+FEATS.

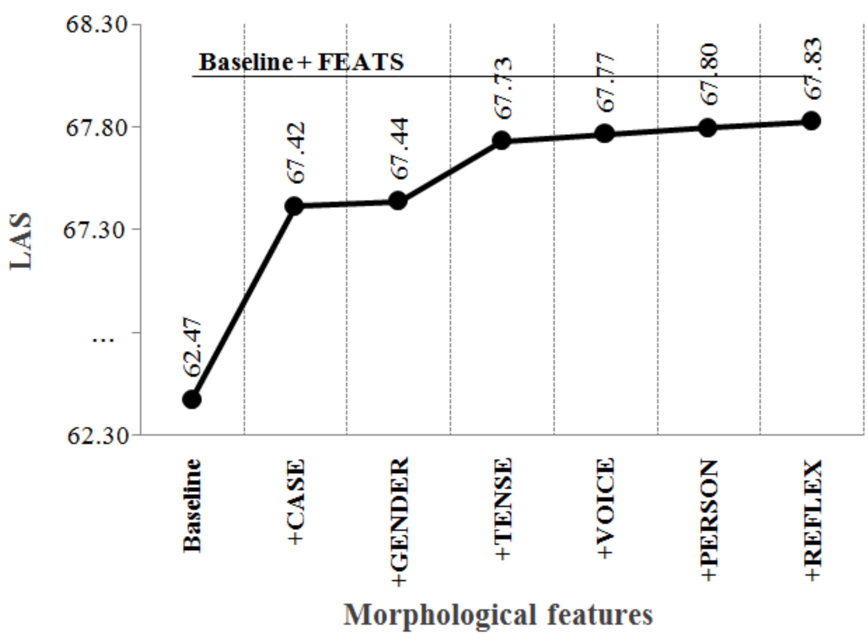

Figure 5: The contribution of individual morphological features in the Predicted condition. The x axis represents feature models incorporating different attributes; the y axis represents LAS. The horizontal line at 68.05 represents the LAS of Baseline+FEATS.

logical analyzer.[8]

It is also worth noting that morphological features have less influence on UAS than LAS, as the gain in UAS over the Baseline is 3-4 percentage points lower compared to LAS. This means that morphology is more important for selecting the type of dependency than for choosing the syntactic head. More precisely, adding morphology improves both recall and precision for the labels SBJ and OBJ, which is probably due primarily to the CASE feature.

Despite the positive effect of morphological information, the best LAS achieved is only 74.66 in the Gold condition and 68.05 in the Predicted condition. An error analysis shows that 38.0% of all LAS errors have an incorrect syntactic head, 12.5% have an incorrect dependency label, and 49.5% have both incorrect. The most commonly occurring problem is the ambiguity between DEP and ROOT dependencies.

For example, in the sentence *atsidūrė Vokietijoje, lankė paskaitas* (he got to Germany, attended lectures) *lankė* (attended) is the dependent of *atsidūrė* (got), because it is the consecutive action performed by the same subject (the subject is expressed implicitly and can be identified according the appropriate verb form). But in the sentence *buvo puiku ir mums, ir jam patiko* (it was great for us and he enjoyed it) *patiko* (enjoyed) is not a dependent of *buvo* (was) but of the root node, because the sentence contains two separate clauses with their subjects and verbs.[9]

Other common ambiguities are among different types of labels that are expressed by the same morphological categories and depends on the context (and the meaning) of the sentence, for example, in the phrase *užželti augalais* (to green with plants), *augalais* (plants) is a dependent of *užželti* (to green) with the OBJ label; in *užsiimti projektais* (to engage in projects) *projektais* (projects) is a dependent of *užsiimti* (to engage) with the MODIF label; and in *pavadinti vardais* (to name with names) *vardais* (names) is a dependent on *pavadinti* (to name) with DEP label. The choice of dependency label in these cases depends on the semantic role of the modifier, corresponding to the question *what* in the first case, the question *how* in the second case, and yet a different relation in the third case. In all these cases morphology does not help to determine the particular label of the dependency relation.

Finally, we note that the results obtained for Lithuanian are in the same range as those reported for Latvian, another Baltic language. Using MaltParser in 10-fold cross-validation on a data set of 2,500 sentences, Pretkalniņa and Rituma (2013) achieve an unlabeled attachment score of 74.6 in the Gold condition and 72.2 in the Predicted conditions, to be compared with 77.8 and 72.8 in our experiments. It should be remembered, however, that the results are not directly comparable due to differences in annotation schemes.

6 Conclusion

In this paper we have presented the first statistical dependency parsing results for Lithuanian. Using the transition-based system MaltParser, we have demonstrated experimentally that the role of morphology is very important for the Lithuanian language. The addition of morphological information resulted in a gain in attachment scores of 7.5 points (labeled) and 3.9 points (unlabeled) with manually validated morphology (the Gold condition) and of 5.6 points (labeled) and 2.5 points (unlabeled) with automatically predicted morphology (the Predicted condition). In the Gold condition, we achieved the best results by adding each morphological feature separately (using the Atom-FEATS representation), but in the Predicted condition adding all features together (using the Set-FEATS representation turned out to be better). The most important morphological feature is CASE, followed by VOICE and TENSE.

Future work includes a more detailed error analysis for the different models, which could throw further light on the impact of different features. It could also be worthwhile to experiment with different feature templates for different morphological categories. For example, for agreement features it seems important to conjoin the values of two words that are candidates for a dependency, while this might not be necessary for features like CASE.

[8] The accuracy is only 86.2%, the lowest of all features.

[9] This type of ambiguity is somewhat artificial, since it arises from the choice to not annotate relations between complete clauses in the Lithuanian Treebank. We expect that parsing accuracy would be improved if all interclausal relations were annotated explicitly.

However, in order to get a major improvement in parsing accuracy, we probably need larger amounts of syntactically annotated data as well as more consistent annotations of interclausal relations.

Acknowledgments

This research is funded by European Union Structural Funds Project “Postdoctoral Fellowship Implementation in Lithuania” (No. VP1-3.1-ŠMM-01) and was initiated when the first author was visiting the Department of Linguistics and Philology at Uppsala University, Sweden.

References

Bharat Ram Ambati, Samar Husain, Joakim Nivre, and Rajeev Sangal. 2010. On the role of morphosyntactic features in hindi dependency parsing. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, SPMRL ’10, pages 94–102, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sabine Buchholz and Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL)*, pages 149–164.

Vidas Daudaravičius, Erika Rimkutė, and Andrius Utka. 2007. Morphological annotation of the Lithuanian corpus. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies (ACL’07)*, pages 94–99.

Gülsen Eryigit, Joakim Nivre, and Kemal Oflazer. 2008. Dependency parsing of Turkish. *Computational Linguistics*, 34.

R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Marija Gimbutas. 1963. *The Balts*. Thames and Hudson.

Yoav Goldberg and Michael Elhadad. 2010. Easy first dependency parsing of modern hebrew. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, SPMRL ’10, pages 103–107, Stroudsburg, PA, USA. Association for Computational Linguistics.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.

Yuval Marton, Nizar Habash, and Owen Rambow. 2013. Dependency parsing of modern standard arabic with lexical and inflectional features. *Computational Linguistics*, 39(1):161–194, March.

Quinn Michael McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.

Joakim Nivre and Jens Nilsson. 2005. Pseudo-projective dependency parsing. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 99–106.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006a. Maltparser: A data-driven parser-generator for dependency parsing. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, pages 2216–2219.

Joakim Nivre, Johan Hall, Jens Nilsson, Gülsen Eryiğit, and Svetoslav Marinov. 2006b. Labeled pseudo-projective dependency parsing with support vector machines. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL)*, pages 221–225.

Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of the CoNLL Shared Task of EMNLP-CoNLL 2007*, pages 915–932.

Joakim Nivre. 2003. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT)*, pages 149–160.

Joakim Nivre. 2008. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34:513–553.

Lauma Pretkalniņa and Laura Rituma. 2013. Statistical syntactic parsing for Latvian. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*, pages 279–289.

Reut Tsarfaty, Djamé Seddah, Yoav Goldberg, Sandra Kuebler, Yannick Versley, Marie Candito, Jennifer Foster, Ines Rehbein, and Lamia Tounsi. 2010. Statistical parsing of morphologically rich languages (spmrl) what, how and whither. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 1–12.

Reut Tsarfaty, Djamé Seddah, Sandra Kübler, and Joakim Nivre. 2013. Parsing morphologicall rich languages: Introduction to the special issue. *Computational Linguistics*, 39:15–22.

Vytautas Zinkevičius. 2000. Lemuoklis – morfologinei analizei [Morphological analysis with Lemuoklis]. *Gudaitis, L. (ed.) Darbai ir dienos*, 24:246–273. (in Lithuanian).