# Special Techniques for Constituent Parsing of Morphologically Rich Languages

**Zsolt Szántó, Richárd Farkas**
University of Szeged
Department of Informatics
{szanto,rfarkas}@inf.u-szeged.hu

## Abstract

We introduce three techniques for improving constituent parsing for morphologically rich languages. We propose a novel approach to automatically find an optimal preterminal set by clustering morphological feature values and we conduct experiments with enhanced lexical models and feature engineering for rerankers. These techniques are specially designed for morphologically rich languages (but they are language-agnostic). We report empirical results on the treebanks of five morphologically rich languages and show a considerable improvement in accuracy and in parsing speed as well.

## 1 Introduction

From the viewpoint of syntactic parsing, the languages of the world are usually categorized according to their level of morphological richness (which is negatively correlated with configurationality). At one end, there is English, a strongly configurational language while there is Hungarian at the other end of the spectrum with rich morphology and free word order (Fraser et al., 2013). A large part of the methodology for syntactic parsing has been developed for English but many other languages of the world are fundamentally different from English. In particular, morphologically rich languages – the other end of the configurational spectrum – convey most sentence-level syntactic information by morphology (i.e. at the word level), not by configuration. Because of these differences the parsing of morphologically rich languages requires techniques that differ from or extend the methodology developed for English (Tsarfaty et al., 2013). In this study, we present three techniques to improve constituent parsing and these special techniques are dedicated to handle the challenges of morphologically rich languages.

Constituency parsers have advanced considerably in the last two decades (Charniak, 2000; Charniak and Johnson, 2005; Petrov et al., 2006; Huang, 2008) boosted by the availability of the Penn Treebank (Marcus et al., 1993). While there is a progress on parsing English (especially the Penn Treebank), the treebanks of morphologically rich languages have been attracted much less attention. For example, a big constituent treebank has been available for Hungarian for almost 10 years (Csendes et al., 2005) and to the best of our knowledge our work is the first one reporting results on this treebank. One reason for the moderate level of interest in constituent parsing of morphologically rich languages is the widely held belief that dependency structures are better suited for representing syntactic analyses for morphologically rich languages than constituent representations because they allow non-projective structures (i.e. discontinuous constituents). From a theoretical point of view, Tsarfaty et al. (2010) point out, however, this is not the same as proving that dependency parsers function better than constituency parsers for parsing morphologically rich languages. For a detailed discussion, please see Fraser et al. (2013).

From an empirical point of view, the organizers of the recent shared task on ‘Statistical Parsing of Morphologically Rich Languages’ (Seddah et al., 2013) provided datasets only for languages having treebanks in both dependency and constituency format and their cross-framework evaluation – employing the unlabeled TedEval (Tsarfaty et al., 2012) as evaluation procedure – revealed that at 4 out of 9 morphologically rich languages, the results of constituent parsers were higher than the scores achieved by the best dependency parsing system. Based on these theoretical issues and empirical results, we support the conclusion of

*Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 135–144,
Gothenburg, Sweden, April 26-30 2014.

Fraser et al. (2013) that "... *there is no clear evidence for preferring dependency parsing over constituency parsing in analyzing languages with rich morphology and instead argue that research in both frameworks is important.*"

In this study, we propose answers to the two main challenges of constituent parsing of morphologically rich languages, which are finding the optimal preterminal set and handling the huge number of wordforms. The size of the preterminal set in the standard context free grammar environment is crucial. If we use only the main POS tags as preterminals, we lose a lot of information encoded in the morphological description of the tokens. On the other hand, using the full morphological description as preterminal yields a set of over a thousand preterminals, which results in data sparsity and performance problems as well. The chief contribution of this work is to propose a novel automatic procedure to find the optimal set of preterminals by **merging morphological feature values**. The main novelties of our approach over previous work are that it is very fast – it operates inside a probabilistic context free grammar (PCFG) instead of using a parser as a black box with re-training for every evaluation of a feature combination – and it can investigate particular morphological feature values instead of removing a feature with all of its values.

Another challenge is that because of the inflectional nature of morphologically rich languages the number of wordforms is much higher compared with English. Hence the number of unknown and very rare tokens – i.e. the tokens that do not appear in the training dataset – is higher here, which hurts the performance of PCFG parsers. Following Goldberg and Elhadad (2013), we **enhance the lexical model** by exploiting an external lexicon. We investigate the applicabilities of fully supervised taggers instead of unsupervised ones for gathering external lexicons.

Lastly, we introduce novel feature templates for an n-best reranker operating on the top of a PCFG parser. These feature templates are **exploiting atomic morphological features** and achieve improvements over the standard feature set engineered for English.

We conducted experiments by the above mentioned three techniques on Basque, French, German, Hebrew and Hungarian, five morphologically rich languages. The BerkeleyParser (Petrov et al., 2006) enriched with these three techniques achieved state-of-the-art results on each language.

## 2 Related Work

Constituent parsing of English is a well researched area. The field has been dominated by data-driven, i.e. treebank-based statistical approaches in the last two decades (Charniak, 2000; Charniak and Johnson, 2005; Petrov et al., 2006). We extend here BerkeleyParser (Petrov et al., 2006), which is a PCFG parser using latent annotations at non-terminals. Its basic idea is to iteratively split each non-terminal into subsymbols thus capturing the different subusage of them instead of manually designed annotations.

The constituent parsing of morphologically rich languages is a much less investigated field. There exist constituent treebanks for several languages along with a very limited number of parsing reports on them. For instance, Petrov (2009) trained BerkeleyParser on Arabic, Bulgarian, French, German and Italian and he reported good accuracies, but there has been previous work on Hebrew (Goldberg and Elhadad, 2013), Korean (Choi et al., 1994) and Spanish (Le Roux et al., 2012) etc. The recently organized 'Statistical Parsing of Morphologically Rich Languages' (Seddah et al., 2013) addressed the dependency and constituency parsing of nine morphologically rich languages and provides useful benchmark datasets for these languages.

Our chief contribution in this paper is a procedure to merge preterminal labels. The related work for this line of research includes the studies on manual refinement of preterminal sets such as Marton et al. (2010) and Le Roux et al. (2012). The most closely related approach to our proposal is Dehdari et al. (2011), who defines metaheuristics to incrementally insert or remove morphological features. Their approach uses parser – training and parsing – as a black box evaluation of a preterminal set. In contrast, our proposal operates as a submodule of the BerkeleyParser, hence does not require the re-training of the parser for every possible preterminal set candidate, thus it is way more faster.

The most successful supervised constituent parsers contain a second feature-rich discriminative parsing step (Charniak and Johnson, 2005; Huang, 2008; Chen and Kit, 2012) as well. At the first stage they apply a PCFG to extract pos-

|  | Basque | French | German | Hebrew | Hungarian |
|---|---|---|---|---|---|
| #sent. in training | 7577 | 14759 | 40472 | 5000 | 8146 |
| #sent. in dev | 948 | 1235 | 5000 | 500 | 1051 |
| #sent. in test | 946 | 2541 | 5000 | 716 | 1009 |
| avg. token/sent. | 12.92 | 30.13 | 17.51 | 25.33 | 21.76 |
| #non-terminal labels | 3000 | 770 | 994 | 1196 | 890 |
| #main POS labels | 16 | 33 | 54 | 46 | 16 |
| unknown token ratio (dev) | 18.35% | 3.22% | 6.34% | 19.94% | 19.94% |

Table 1: Basic statistics of the treebanks used.

sible parses. The *n-best list parsers* keep just the 50-100 best parses according to the PCFG (Charniak and Johnson, 2005). These methods employ a large feature set (usually a few million features) (Collins, 2000; Charniak and Johnson, 2005). These feature sets are engineered for English. In this study, we introduce feature templates for exploiting morphological information and investigate their added value over the standard feature sets.

## 3 Experimental Setup

We conducted experiments on the treebanks of the 2013 shared task on 'Statistical Parsing of Morphologically Rich Languages' (Seddah et al., 2013). We used the train/dev/test splits of the shared task's Basque (Aduriz et al., 2003), French (Abeillé et al., 2003), Hebrew (Sima'an et al., 2001), German (Brants et al., 2002) and Hungarian (Csendes et al., 2005) treebanks. Table 1 shows the basic statistics of these treebanks, for a more detailed description about their annotation schemata, domain, preprocessing etc. please see Seddah et al. (2013).

As evaluation metrics we employ the PARSEVAL score (Abney et al., 1991) along with the exact match accuracy (i.e. the ratio of perfect parse trees). We use the evalb implementation of the shared task[1].

## 4 Enhanced Lexical Models

Before introducing our proposal and experiments with preterminal set optimisation, we have to offer a solution for the out-of-vocabulary (OOV) problem, which – because of the inflectional nature – is a crucial problem in morphologically rich languages. We follow here Goldberg and Elhadad (2013) and enhance a lexicon model trained on the training set of the treebank with frequency information about the possible morphological analyses of tokens. We estimate the tagging probability $P(t|w)$ of the tag $t$ given the word $w$ by

$$P(t|w) = \begin{cases} P_{tb}(t|w), & \text{if } c(w) \geq K \\ \frac{c(w)P_{tb}(t|w)+P_{ex}(t|w)}{1+c(w)}, & \text{otherwise} \end{cases}$$

where $c(w)$ is the count of $w$ in the training set, $K$ is predefined constant, $P_{tb}(t|w)$ is the probability estimate from the treebank (the relative frequency with smoothing) and $P_{ex}(t|w)$ is the probability estimate from an external lexicon. We calculate the emission probabilities $P(w|t)$ from the tagging probabilities $P(t|w)$ by applying the Bayesian rule.

The key question here is how to construct the external lexicon. For a baseline, Goldberg and Elhadad (2013) suggest using the uniform distribution over all possible morphological analyses coming from a morphological analyser ('uniform').

Goldberg and Elhadad (2013) also report considerable improvements over the 'uniform' baseline by relative frequencies counted on a large corpus which was automatically annotated in the unsupervised POS tagging paradigm (Goldberg et al., 2008). Here we show that even a supervised morphological tagger without a morphological analyzer can achieve the same level of improvement. We employ MarMot[2] (Mueller et al., 2013) for predicting full morphological analysis (i.e. POS tags and morphological features jointly). MarMot is a Conditional Random Field tagger which incrementally creates forward-backward lattices of increasing order to prune the

[1] Available at `http://pauillac.inria.fr/~seddah/evalb_spmrl2013.tar.gz`. An important change in this version compared to the original evalb is the penalization of unparsed sentences.

[2] `https://code.google.com/p/cistern/`

sizable space of possible morphological analyses. We used MarMoT with the default parameters. This purely data-driven tagger achieves a tagging accuracy of 97.6 evaluated at full morphological analyses on the development set of the Hungarian treebank, which is competitive with the state-of-the-art Hungarian taggers which employ language-specific rules (e.g. magyarlanc (Zsibrita et al., 2013)). The chief advantage of using MarMot instead of an unsupervised tagger is that the former does not require any morphological lexicon/analyser (which can lists the possible tags for a given word). This morphological lexicon/analyser is language-dependent, usually handcrafted and it has to be compatible with the treebank in question. In contrast, a supervised morphological tagger can build a reasonable tagging model on the training part of the treebanks – especially for morphologically rich languages, where the tag ambiguity is generally low – thus each of these problems is avoided.

Table 2 shows the results of various $P_{ex}(t|w)$ estimates on the Hungarian development set. The first row 'BerkeleyParser' is our absolute baseline, i.e. the original implementation of BerkeleyParser[3] defining signatures for OOVs. For the 'uniform' results, we used the morphological analyser module of magyarlanc (Zsibrita et al., 2013). The last two rows show the results achieved by training MarMot on the treebank's training dataset, having tagged the development set plus a huge unlabeled corpus (10M sentences from the Hungarian National Corpus (Váradi, 2002)) with it then having counted relative tag frequencies. We report scores on only using the frequencies from the development set ('dev') and from the concatenation of the development set and the huge corpus ('huge').

After a few preliminary experiments, we set $K = 7$ and use this value thereafter.

Table 2 shows that even 'dev' yields a considerable improvement over the baseline parser and 'uniform'. These results are also in line with the findings of Goldberg and Elhadad (2013), i.e. 'uniform' has some added value and using relative frequencies gathered from automatically tagged corpora contributes more. Although we can see another nice improvement by exploiting unlabeled corpora ('huge'), we will use the 'dev' setting in the experiments of the next sections as we did not have access to huge, in-domain unlabeled corpora for each language used in this study.

| | PARSEVAL | EX |
|---|---|---|
| BerkeleyParser | 87.22 | 12.75 |
| uniform | 87.31 | 14.78 |
| dev | 88.29 | 15.22 |
| huge | 89.27 | 16.97 |

Table 2: The results achieved by using various external lexical models on the Hungarian development set.

## 5 Morphological Feature Values as Preterminals

Finding the optimal set of morphological features incorporating into the perterminal labels is crucial for any PCFG parsers. Removing morphological features might reduce data sparsity problems while it might lead to loss of information for the syntactic parser. In this section, we propose a novel method for automatically finding the optimal set of preterminals then we present empirical results with this method and compare it to various baselines.

**Merge Procedure for Morphological Feature Values:** There have been studies published on the automatic reduction of the set of preterminals for constituent parsing. For instance, Dehdari et al. (2011) proposed a system which iteratively removes morphological features as a unit then evaluates the preterminal sets by running the training and parsing steps of a black-box constituent parser. Our motivation here is two-fold. First, morphological features should not be handled as a unit because different values of a feature might behave differently. Take for instance the degree feature in Hungarian adjectives. Here the values positive and superlative behave similarly (can be merged) while distinguishing comparative and positive+superlative is useful for syntactic parsing because comparative adjectives often have an argument (e.g. x is more beautiful than y) while positive and superlative adjectives are not syntactic governors thus have no arguments. Second, keeping a morphological feature can be useful for particular POS tags and useless at other particular POS tags (e.g. the number of possessed in Hungarian for nouns and pronouns).

[3] http://code.google.com/p/berkeleyparser/

**Algorithm 1** The preterminal set merger algorithm.

```
1. training the standard BerkeleyParser using only main POS tags as preterminals
2. merging each subsymbols at the preterminal level
3. for each POS tag - morphological feature pair
   (a) split the POS tag for the values of the morphological feature[4]
   (b) recalculating the rule probabilities where there are preterminals in the right-hand side by uniformly distribute the probability mass among subsymbols
   (c) set the lexical probabilities according to the relative frequencies of morphological values counted on gold standard morphological tags of the treebank
   (d) running 10 iterations of the Expectation-Maximization procedure on the whole treebank initialized with (b)-(c)
   (e) constructing a fully connected graph whose nodes are the morphological values of the feature in question
   (f) for every edge of the graph, calculate the loss in likehood for the merging the two subsymbols (the same way as for BerkeleyParser's merge procedure)
4. removing edges from the entire set of graphs (controlled by the parameter $th$)
5. merge the morphological values of the graphs' connected components
```

Based on these observations we propose a procedure which starts from the full morphological description of a treebank then iteratively merges particular morphological feature values and it handles the same feature at the different POS tags separately. The result of this procedure is a clustering of the possible values of each morphological feature. The removal of a morphological feature is a special case of our approach because if the values of the feature in question form one single cluster it does not have any discriminative function anymore. Hence our proposal can be regarded as a generalisation of the previous approaches.

This general approach requires much more evaluation of intermediate candidate preterminal sets, which is not feasible within the external black-box parser evaluation scenario (training and parsing an average sized treebank by the BerkeleyParser takes more than 1 hour). Our idea here is that retraining a parser for the evaluation of each preterminal set candidates is not necessary. They key objective here is to select among preterminal sets based on their usefulness for the syntactic parser. This is the motivation of the merge procedure of the BerkeleyParser. After randomly splitting nonterminals, BerkeleyParser calculates for each split the loss in likelihood incurred when merging the subsymbols back. If this loss is small, the new annotation does not carry enough useful information and can be removed (Petrov et al., 2006). Our task is the same at the preterminal level. Hence at the preterminal level, – instead of using the automatic subsymbol splits of the BerkeleyParser – we call this merging procedure over the morphological feature values. Algorithm 1 shows our proposal for the preterminal merging procedure.

**Baseline Preterminal Set Constructions:** The two basic approaches for preterminal set construction are the use of only the main POS tag set ('mainPOS') and the use of the full morphological description as preterminals ('full'). For Hungarian, we also had access to a linguistically motivated, hand-crafted preterminal set ('manual') which was designed for a morphological tagger (Zsibrita et al., 2013). This manual code set keeps different morphological features at different POS tags and merges morphological values instead of fully removing features hence it inspired our automatic merge procedure introduced in the previous section.

Our last baseline is the repetition of the experiments of Dehdari et al. (2011). For this, we started from the full morphological feature set and completely removed features (from all POS) one-by-one then re-trained our parser. We observed the greatest drop in PARSEVAL score at removing the

| | Basque | French | German | Hebrew | Hungarian |
|---|---|---|---|---|---|
| mainPOS | 68.8/3.9 16 | 78.4/13.9 33 | 82.3/38.7 54 | 88.3/12.0 46 | 82.6/7.3 16 |
| full | **81.8/18.4** 2976 | 78.9/15.0 676 | **82.3/40.3** 686 | 88.9/**15.2** 257 | 88.3/15.2 680 |
| preterminal merger | 81.6/16.9 2791 | **79.7/15.6** 480 | 82.3/39.3 111 | **89.0**/14.6 181 | **88.5/15.4** 642 |

Table 3: PARSEVAL / exact match scores on the development sets. The third small numbers in cells show the size of the preterminal sets.

'Num' feature and the least severe one at removing 'Form'. 'Num' denotes number for verbs and nominal elements (nouns, adjectives and numerals), and since subject-verb agreement is determined by the number and person features of the predicate (the verb) and the subject (the noun), deleting the feature 'Num' results in a serious decline in performance. On the other hand, 'Form' denotes whether a conjunction is single or compound (which is a lexical feature) or whether a number is spelt with letters, Arabic or Roman numbers (which is an orthographic feature). It is interesting to see that their deletion hardly harms the PARSEVAL scores, moreover, it can even improve the exact match scores, which is probably due to the fact that the distinction between different orthographic versions of the same number (e.g. *6* and *VI*) just confused the parser. On the other hand, members of a compound conjunction are not attached to each other in any way in the parse tree, and behave similar to single compounds, so this distinction might also be problematic for parsing.

**Results with Various Preterminal Sets:** Table 4 summarizes the results achieved by our four baseline methods along with the scores of two preterminal sets output by our merger approach at two different merging threshold $th$ value.

| | #pt | PARSEVAL | EX |
|---|---|---|---|
| mainPOS | 16 | 82.36 | 5.52 |
| manual | 72 | 85.38 | 9.23 |
| full | 680 | 88.29 | 15.22 |
| full - Num | 479 | 87.43 | 14.49 |
| full - Form | 635 | 88.24 | 15.73 |
| merged ($th = 0.5$) | 378 | 88.36 | **15.92** |
| merged ($th = 0.1$) | 642 | **88.52** | 15.44 |

Table 4: The results achieved by using various preterminal sets on the Hungarian development set.

The difference between mainPOS and full is surprisingly high, which indicates that the morphological information carried in preterminals is extremely important for the constituent parser and the BerkeleyParser can handle preterminal sets of the size of several hundreds. For Hungarian, we found that the full removal of any feature cannot increase the results. This finding is contradictory with Dehdari et al. (2011) in Arabic, where removing 'Case' yielded a gain of 1.0 in PARSEVAL. We note that baselines for Arabic and Hungarian are also totally different, Dehdari et al. (2011) reports basically no difference between mainPOS and full in Arabic.

We report results of our proposed procedure with two different merging thresholds. The $th = 0.1$ case merges only a few morphological feature values and it can slightly outperform the 'full' setting (statistically significant[5] in exact match.). On the other hand, the $th = 0.5$ setting is competitive with the 'full' setting in terms of parsing accuracy but it uses only the third of the preterminals used by 'full'. Although it is not statistically better than 'full' in accuracy, it almost halves the running time of parsing[6].

Table 3 summarizes the results achieved by the most important baselines and our approach along with the size of the particular preterminal sets applied. The 'full' results outperform 'mainPOS' at each language with a striking difference at Basque and Hungarian. These results show that – contradictory to the general belief – the detailed morphological description is definitely useful in constituent parsing as well. The last row of the table contains the result achieved by our merger approach. Here we run experiments with several merging threshold $th$ values and show the highest scores for each language.

Our merging proposal could find a better preterminal set than full on French and Hungarian, it found a competitive tag set in terms of accuracies

[5] According to two sample t-test with $p<0.001$.

[6] Parsing the 1051 sentences of the Hungarian development set takes 15 and 9 minutes with full and $th = 0.5$ respectively (on an Intel Xeon E7 2GHz).

which are much smaller than full on German and Hebrew and it could not find any useful merge at Basque. The output of the merger procedure consists of one sixth of preterminals compared with full. Manually investigating the clusters, we can see that it basically merged every morphological feature except case at nouns and adjectives (but merged case at personal pronouns). This finding is in line with the experimental results of Fraser et al. (2013).

## 6 Morphology-based Features in n-best Reranking

$n$-best rerankers (Collins, 2000; Charniak and Johnson, 2005) are used as second stage after a PCFG parser and they usually achieve considerable improvement over the first stage parser. They extract a large feature set to describe the $n$ best output of a PCFG parser and they select the best parse from this set (i.e. rerank the parses). Here, we define feature templates exploiting morphological information and investigate their added value for the standard feature sets (engineered for English). We reimplemented the feature templates from Charniak and Johnson (2005) and Versley and Rehbein (2009) excluding the features based on external corpora and use them as our baseline feature set.

We used $n = 50$ in our experiment and followed a 5-fold-cross-parsing (a.k.a. jackknifing) approach for generating unseen parse candidates for the training sentences (Charniak and Johnson, 2005). The reranker is trained for the maximum entropy objective function of Charniak and Johnson (2005), i.e. the sum of posterior probabilities of the oracles. We used a slightly modified version of the Mallet toolkit for reranking (McCallum, 2002) and L2 regularizer with its default value for coefficient.

The feature templates of the baseline feature set frequently incorporate preterminals as atomic feature. As a first step, we investigated which preterminal set is the most useful for the baseline feature set. We took the 50 best output from the parser using the merged preterminal set and used its preterminals ('merged') or only the main POS tag ('mainPOS') as atomic building blocks for the reranker's feature extractor. Table 5 shows that mainPOS outperformed full. This is probably due to data sparsity problems.

Based on this observation, we decided to use mainPOS as preterminal in the atomic building block of the baseline features and designed new feature templates capturing the information in the morphological analysis. We experimented with the following templates:

For each preterminal of the candidate parse and for each morphological feature value inside the preterminal we add the pair of wordform and morphological feature value as a new feature. In a similar way, we define a reranker feature from every morphological feature value of the head word of the constituent. For each head-daughter attachment in the candidate parse we add each pair of the morphological feature values from the head words of the attachment's participants. Similarly we take each combination of head word's morphological features values from sister constituents.

The first two templates enable the reranker to incorporate information into its learnt model from the rich morphology of the language at the lexical and constituent levels, while the last two templates might capture (dis)agreement at the morphological level. The motivation for using these features is that because of the free(er) word order of morphologically rich languages, morphological (dis)agreement can be a good indicator of attachment.

Table 5 shows the added value of these feature templates over mainPOS ('extended'), which is again statistically significant in exact match. Exploiting the morphological agreement in syntactic parsing has been investigated in previous studies, e.g. the Bohnet parser (Bohnet, 2010) employs morphological feature value pairs similar to our feature templates and Seeker and Kuhn (2013) introduces an integer linear programming framework including constraints for morphological agreement. However, these works focus on dependency parsing and to the best of our knowledge, this is the first study on experimenting with atomic morphological features and their agreement in a constituency parsing.

| | PARSEVAL | EX |
|---|---|---|
| reranker (merged morph) | 89.05 | 18.45 |
| reranker (mainPOS) | 89.33 | 18.64 |
| reranker (extended) | **89.47** | **20.35** |

Table 5: The results achieved by using various feature template sets for 50-best reranking on the Hungarian development set.

| | Basque | French | German | Hebrew | Hungarian |
|---|---|---|---|---|---|
| BerkeleyParser | 79.21 / 19.03 | 79.53 / 18.46 | 74.77 / 26.56 | 87.87 / 14.53 | 88.22 / 26.96 |
| + Lexical model | 82.02 / 25.69 | 78.91 / 17.87 | 75.64 / 28.36 | 88.53 / 13.69 | 89.09 / 26.76 |
| + Preterminal merger | 83.19 / 24.74 | 79.53 / 18.58 | 77.12 / **30.02** | 88.07 / 13.83 | 89.15 / 28.05 |
| + reranker | 83.81 / 25.66 | 80.31 / 18.91 | **77.78** / 29.80 | 88.38 / 15.12 | 89.57 / 30.23 |
| + reranker + morph feat | **84.03 / 26.28** | **80.41 / 20.07** | 77.74 / 29.23 | **88.55 / 15.24** | **89.91 / 30.55** |

Table 6: PARSEVAL / exact match scores on the test sets.

## 7 Results of the Full System

After our investigations focusing on building blocks of our system independently from each other on the development set, we parsed the test sets of the treebanks adding steps one-by-one. Table 6 summarizes our final results. We start from the BerkeleyParser using the full morphological descriptions as preterminal set, then we enrich the lexical model with tagging frequencies gathered from the automatic parsing of the test sets ('+ lexical model'). In the third step we replace the full preterminal set by the output of our preterminal merger procedure ('+ preterminal merger'). We tuned the merging threshold of our method on the development set for each language. The last two rows contain the results achieved by the 50-best reranker with the standard feature set ('+ reranker') and with the feature set extended by morphological features ('+ morph features').

The enhanced lexical model contributes a lot at Basque and considerable improvements are present at German and Hungarian as well while it harmed the results in French. The advance of the preterminal merger approach over the full setting is clear at French and Hungarian, similarly to the development set. It is interesting that an rationalized preterminal set could compensate the loss suffered by a inadequate lexical model at French.

Although the reranking step could further improve the results at each languages we have to note that the gain (0.5 in average) is much smaller here than the gains reported on English (over 1.5). This might be because of the high number of wordforms at morphologically rich languages i.e. most of feature templates are incorporate the words itself and the huge dictionary can indicate data sparsity problems again. Our morphology-based reranking features yielded a moderate improvement at four languages, but we believe there a lots of space for improvement here.

## 8 Conclusions

In this study we introduced three techniques for better constituent parsing of morphologically rich languages. We believe that research in constituency parsing is important next to dependency parsing. In general, we report state-of-the-art results with constituent parsers with our entirely language-agnostic techniques.

Our chief contribution here is the preterminal merger procedure. This is a more general approach than previous proposals and still much faster thank to operating on probabilities from a PCFG instead of employing a full train+parse step for evaluating every preterminal set candidate. We found that the inclusion of the rich morphological description into the preterminal level is crucial for parsing morphologically rich languages. Our proposed preterminal merger approach could outperform the full setting at 2 out of 5 languages, i.e. we have reported gains in parsing accuracies by merging morphological feature values. At the other languages, the results with the full preterminal set and our approach are competitive in terms of parsing accuracies while our approach could achieve these scores with a smaller preterminal set, which leads to considerable parsing time advantages.

We also experimented with exploiting external corpora in the lexical model. Here we showed that automatic tagging of an off-the-shelf supervised morphological tagger (MarMot) can contribute to the results. Our last experiment was carried out with the feature set of an $n$-best reranker. We showed that incorporating feature templates built on morphological information improves the results.

## Acknowledgments

This work was supported in part by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TÁMOP-4.2.2.C-11/1/KONV-2012-0013).

## References

Anne Abeillé, Lionel Clément, and François Toussenel. 2003. Building a treebank for french. In Anne Abeillé, editor, *Treebanks*. Kluwer, Dordrecht.

S. Abney, S. Flickenger, C. Gdaniec, C. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini, and T. Strzalkowski. 1991. Procedure for quantitatively comparing the syntactic coverage of english grammars. In E. Black, editor, *Proceedings of the workshop on Speech and Natural Language*, pages 306–311.

I. Aduriz, M. J. Aranzabe, J. M. Arriola, A. Atutxa, A. Díaz de Ilarraza, A. Garmendia, and M. Oronoz. 2003. Construction of a Basque dependency treebank. In *TLT-03*, pages 201–204.

Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 89–97.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER treebank. In Erhard Hinrichs and Kiril Simov, editors, *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT 2002)*, pages 24–41.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 173–180.

Eugene Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*, pages 132–139.

Xiao Chen and Chunyu Kit. 2012. Higher-order constituent parsing and parser combination. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1–5.

Key-Sun Choi, Young S Han, Young G Han, and Oh W Kwon. 1994. Kaist tree bank project for korean: Present and future development. In *Proceedings of the International Workshop on Sharable Natural Language Resources*, pages 7–14. Citeseer.

Michael Collins. 2000. Discriminative reranking for natural language parsing. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 175–182.

Dóra Csendes, János Csirik, Tibor Gyimóthy, and András Kocsor. 2005. The Szeged Treebank. In *TSD*, pages 123–131.

Jon Dehdari, Lamia Tounsi, and Josef van Genabith. 2011. Morphological features for parsing morphologically-rich languages: A case of arabic. In *Proceedings of the Second Workshop on Statistical Parsing of Morphologically Rich Languages*, pages 12–21, Dublin, Ireland, October. Association for Computational Linguistics.

Alexander Fraser, Helmut Schmid, Richárd Farkas, Renjing Wang, and Hinrich Schütze. 2013. Knowledge sources for constituent parsing of german, a morphologically rich and less-configurational language. *Computational Linguistics*, 39(1):57–85.

Yoav Goldberg and Michael Elhadad. 2013. Word segmentation, unknown-word resolution, and morphological agreement in a hebrew parsing system. *Computational Linguistics*, 39(1):121–160.

Yoav Goldberg, Meni Adler, and Michael Elhadad. 2008. EM can find pretty good HMM POS-taggers (when given a good start). In *Proceedings of ACL-08: HLT*, pages 746–754.

Liang Huang. 2008. Forest reranking: Discriminative parsing with non-local features. In *Proceedings of ACL-08: HLT*, pages 586–594.

Joseph Le Roux, Benoit Sagot, and Djamé Seddah. 2012. Statistical parsing of spanish and data driven lemmatization. In *Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, pages 55–61.

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

Yuval Marton, Nizar Habash, and Owen Rambow. 2010. Improving arabic dependency parsing with lexical and inflectional morphological features. In *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 13–21.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.

Thomas Mueller, Helmut Schmid, and Hinrich Schütze. 2013. Efficient higher-order CRFs for morphological tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 322–332.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440.

Slav Petrov. 2009. *Coarse-to-Fine Natural Language Processing*. Ph.D. thesis, University of California at Bekeley, Berkeley, CA, USA.

Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho D. Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola Galletebeitia, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Yuval Marton, Joakim Nivre, Adam Przepiórkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński, and Alina Wróblewska. 2013. Overview of the SPMRL 2013 shared task: A cross-framework evaluation of parsing morphologically rich languages. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 146–182.

Wolfgang Seeker and Jonas Kuhn. 2013. Morphological and syntactic case in statistical dependency parsing. *Computational Linguistics*, 39(1):23–55.

Khalil Sima'an, Alon Itai, Yoad Winter, Alon Altman, and Noa Nativ. 2001. Building a Tree-Bank for Modern Hebrew Text. In *Traitement Automatique des Langues*.

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, Joakim Nivre, and Evelina Andersson. 2012. Cross-framework evaluation for statistical parsing. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 44–54.

Reut Tsarfaty, Djamé Seddah, Sandra Kübler, and Joakim Nivre. 2013. Parsing morphologically rich languages: Introduction to the special issue. *Computational Linguistics*, 39(1):15–22.

Tamás Váradi. 2002. The hungarian national corpus. In *In Proceedings of the Second International Conference on Language Resources and Evaluation*, pages 385–389.

Yannick Versley and Ines Rehbein. 2009. Scalable discriminative parsing for german. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, pages 134–137.

János Zsibrita, Veronika Vincze, and Richárd Farkas. 2013. magyarlanc: A toolkit for morphological and dependency parsing of hungarian. In *Proceedings of RANLP*.