Constituency Parse Reranking for Morphologically Rich Languages

Zsolt Szántó, Richárd Farkas

University of Szeged, Department of Informatics
Árpád tér 2, H-6700 Szeged, Hungary
szantozs@inf.u-szeged.hu, rfarkas@inf.u-szeged.hu

Abstract: In this article we introduce a constituent parsing system which can achieve state-of-the-art results on morphologically rich languages. Our system consists of a Probabilistic Context Free Grammars (PCFG) and n best reranking steps. We compare two methods to handle lexical sparsity in a PCFG parser. The n best reranking step, the discriminative reranker extracts large amount of features from n best parses of the PCFG parser and selects the best tree from these parses. We introduce three feature templates which extend the standard feature set of rerankers. We propose to extract features from Brown clustering – which is a context-based clustering over the words – and analyze the effect of dependency-based and morphology-based feature templates. The effects of these techniques are evaluated on datasets of eight morphologically rich languages.

Keywords: syntactic parsing; constituent parsing; morphologically rich languages; lexical sparsity; Brown clustering

1 Introduction

Syntax, in natural languages, describes the structure of a sentence and the grammatical relations between the words. In computational linguistics, syntactic parsing generally supports higher level tasks since the knowledge of the syntactic structure of a sentence can contribute to many natural language processing end-user applications like, machine translation and information retrieval. In this paper, we focus on constituent parsing, which is one of the most commonly used syntactic representations in computational linguistics (see Figure 1 for an example of constituent parse tree).

In the beginning, like many other fields of natural language processing, syntactic parsing systems focused on English. English is a strongly configurational language, while some morphologically rich languages (like Hungarian) exhibit free word order and express grammatical roles with morphology whereas in English, they are expressed by word order. Another difference between English

and morphologically rich languages is the number of word forms. Because of the inflectional nature of morphologically rich languages they contain many more word forms. The huge number of word forms is the reason for the lexical sparsity issue, which makes the parsing task more difficult. Here, we comparatively evaluate two methods targeting lexical sparsity.

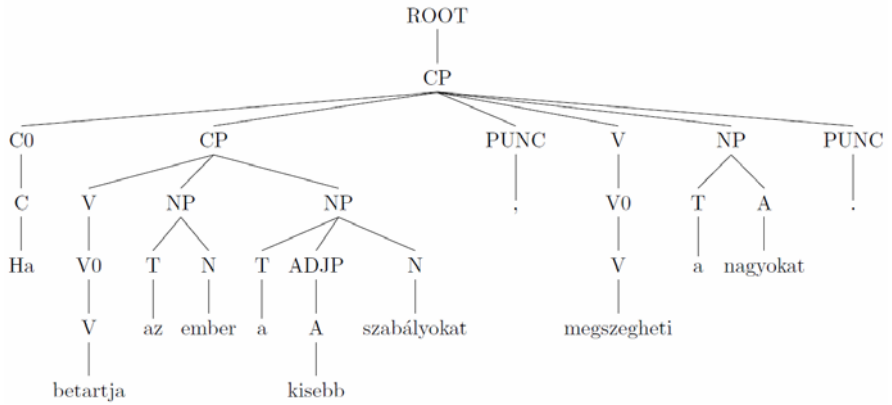

Figure 1
Constituency tree of a Hungarian sentence from Orwell's 1984 ('*If you kept the small rules, you could break the big ones.*')

Constituent parsing systems usually use a discriminative n-best reranking step. These reranking systems can improve the performance of a first-stage PCFG parsing methods [20]. The first-stage parser must be fast so as to be able to select the best trees from all the possible parses, and the reranker can extract a rich feature set to describe the *n* best parses of the original parser and rerank the *n* candidate trees according to the features utilizing machine learning methods [6].

In this article we focus on this reranking step. On the Statistical Parsing of Morphologically Rich Languages (SPMRL) Shared Task datasets [22] – consisting of training and evaluation datasets for morphologically rich languages – we introduce and evaluate in detail three different feature sets which can improve the results in the case of these morphological rich languages. One of these feature sets directly focuses on the morphological rich data and exploits atomic morphological features, another applies information from dependency parsing, and the last one is based on Brown clusters. The latter method groups the words to hierarchical categories based on their context. This method also can help in the case of the out-of-vocabulary issue, which is the consequence of the large number of word forms.

The main contribution of this article is the introduction and the evaluation of these feature sets in the discriminative reranking step.

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 [5, 6, 20]. We extend here the BerkeleyParser [20], 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. We use two different methods to handle the huge number of word forms. First, we create an extended lexicon [24] for the better estimation of tagging probabilities. Second, we use another method where we replace the rare words with their predicated POS tags [2]. Petrov [19] showed that the product of different grammars can improve the accuracy.

The most successful supervised constituent parsers contain a second feature-rich discriminative parsing step [6, 7, 14] as well. At the first stage they apply a PCFG to extract possible parses. The *n-best list rerankers* keep just the 50-100 best parses according to the PCFG [6]. These methods employ a large feature set (usually a few million features) describing parse candidates then use supervised machine learning techniques (learning-to-rank) to select the best parser based on the features [6, 10]. These feature sets are also engineered for English. To the best of our knowledge there is only one previous work which deals with the reranking of morphologically rich languages [12].

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, [18] trained BerkeleyParser on Arabic, Bulgarian, French, German and Italian and he reported good accuracies, and there has been previous work on Hebrew [13], Korean [8] and Spanish [15] etc. The 'Statistical Parsing of Morphologically Rich Languages' [21] addressed the dependency and constituency parsing of morphologically rich languages and provides useful benchmark datasets for these languages.

Our chief contribution in this paper is the introduction of three feature sets for morphologically rich languages in the second stage reranking. Previously, the dependency based features were successfully applied to German [11]. Here, we experiment with them on seven morphologically rich languages. The morphological features [24] were designed especially for morphologically rich languages. To the best of our knowledge, the Brown clustering [4] based features have been previously not used in the context of reranking.

3 Experimental Setup

In our experiments we used the SPMRL 2014 Shared Task's data. It contains constituent and dependency trees on morphologically rich languages. This shared task also contains large unlabeled data sets in every language. We use seven languages from the shared task's data (Basque, French, German, Hebrew, Hungarian, Polish, and Swedish). Table 1 contains the basic statistics of the treebanks.

Table 1
Basic statistics of the treebanks used

	baq	fra	ger	heb	hun	pol	swe
#sent. in training	7577	14759	40472	5000	8146	6578	5000
#sent. in dev	948	1235	5000	500	1051	821	494
avg. token/sent.	12.93	30.13	17.51	25.34	21.75	10.15	15.59
#non-terminal labels	2764	693	709	507	676	675	137
#main POS labels	49	33	56	52	16	29	25
unk. token ratio (dev)	18.35	3.23	7.64	9.65	19.94	24.76	11.88

We use the PARSEVAL [1] score, which is the most common metric for the evaluation of constituent parsing. It is a standard F-score over labeled bracketings.

4 The Constituency System for Morphologically Rich Languages

The constituent parsing systems usually use two steps. The first step is a PCFG parser which selects the best parses from all possible trees relatively fast. In the first step we employed here the Berkeley Parser [20]. We introduce the second step in the Section 4.4.

4.1 Lexical Sparsity

Like many other parsers, the Berkeley Parser [20] was designed to English, but the huge number of wordforms create new challenges in the case of morphologically rich languages. We used two fundamentally different methods to handle the out-of-vocabulary issue. We built two standalone parsing systems based on these methods. One of these two techniques is the usage of extended lexicons [24]. We followed [13] and enhanced a lexicon model trained on the training set of the treebank with frequency information about the possible morphological analyses of tokens (*ExtendLex*). We estimated the tagging probability $P(t|w)$ of the tag t given the word w by

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

where $c(w)$ is the count of w in the training set, K is a 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 calculated the emission probabilities $P(w|t)$ from the tagging probabilities $P(t|w)$ by applying the Bayesian rule. This method can exploit the available unlabeled data [24].

The other method is based on Clark and Curran's [9] work, where they replaced the rare words with their predicted POS tags in CCG grammars. We also used this strategy in our constituent parsing framework (*Replace*).

Both methods require automatically annotated POS tags. *Replace* uses these for replacing the rare words and *ExtendLex* for calculating new tagging probabilities. To get this information, we used MarMoT [17], which is a language independent POS-tagger. We trained MarMoT on the SPMRL's training sets. In some languages (Basque, French, German, Hungarian, Polish) we analyzed the word forms with the language-specific morphological analyzer and we used this information as features in MarMoT. Table 2 contains the accuracy of the part of speech tagging.

Table 2
POS and morphological feature accuracies of part of speech tagging on the development sets

	baq	fra	ger	heb	hun	pol	swe
MarMoT POS	97.52	97.08	97.98	96.97	98.49	98.39	97.40
MarMoT morph	87.81	89.36	90.38	97.15	97.45	91.00	97.16

4.2 Preterminal Set

The selection of the preterminal set in a PCFG is crucial and challenging in morphologically rich languages since they can consist of thousands of labels. To handle this issue, we have several options to use the morphological information at the preterminal level. First, we can use the full morphological descriptions (for example, *N##SubPOS=c|Num=s|Cas=n* means that this is a common noun with nominative case in singular) or second, just the main part of speech tags (for example *N*). The full morphological description provides more information, but increases the data sparsity problem. The selection of the labels at the preterminal level is crucial to PCFG parsers, hence we can get very different results with different preterminal level [24].

Table 3
PARSEVAL scores on the development sets

	baq	fra	ger	heb	hun	pol	swe
Berkeley mainPOS	72.32	79.35	82.26	88.71	83.84	86.75	75.19
Berkeley fullMorph	77.82	79.17	80.22	88.40	87.18	85.06	72.82

Table 3 contains the results of the Berkeley Parser with main POS and full morphological description. We can see that in Basque and Hungarian the usage of full morphological description obtains higher results, but in the other languages we can get better results with main POS. Our two methods used the POS tags differently. In the case of *ExtendLex* we used the full morphological description at the preterminal level. In the case of *Replace* the preterminal level always contained the main POS tags.

4.3 Product Parser

The Berkeley Parser uses an iterative expectation–maximization approach to calculate the latent variables [20]. This is a local search algorithm and the result of this optimization is highly dependent on the initialization of the variables. In every iteration the Berkeley Parser splits the latent variables in two with a small amount of randomness. Petrov showed [19] that the modification of these random values can change the result of the parsing. We can get different parsers when we only change the random seed. If we product the probabilities of the same sentence with these different grammars, we can get better scores than in the case when we just used one grammar. Based on this experience, we trained 8 grammars with different random seeds and we obtained the product of them.

4.4 Reranker for Morphologically Rich Languages

The second step of our constituency pipeline is discriminative reranking. We conducted ranking experiments on the 50-best outputs of the product grammars. We used a slightly modified version of the Mallet toolkit [16], where the reranker was trained for the maximum entropy objective function of Charniak and Johnson [6] and used the standard feature set from Charniak and Johnson [6] and Collins [10]. We propose here new feature templates exploiting automatic dependency parses of the sentence in question; Brown clusters; and atomic morphological feature values. Our purpose here is to investigate the efficiency of these feature templates in morphologically rich languages. For these studies we used the product grammar configuration. Here, we present our feature templates in more detail.

4.4.1 Morphology-based Features

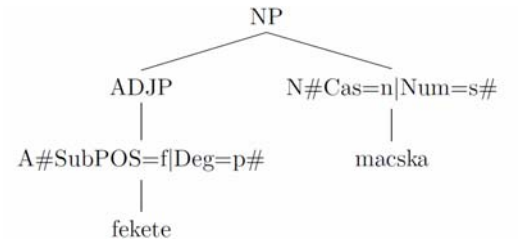

Figure 2
Constituency tree of a Hungarian noun phrase (*black cat*)

We created features from the full morphological description by using each morphological feature separately [24]. This approach allows us to combine a word with its morphological features, for example the word *macska*, which is a noun in nominative case, we made the (macska-N-Cas=n) feature (see Figure 2).

We also made features from the phrases, where we used the combination of a phrase and their head's morphological features. In our example we can create two features from the *NP*: NP-N-Cas=n and NP-N-Num=s.

If the phrase contains other children, we can create more features from these parent-daughter relations. The new features contain the morphological features of the head of the parent and the head of the daughter. In the *fekete macska* example we can get 4 features: NP-Cas=n-ADJP-SubPOS=f, NP-Cas=n-ADJP-Deg=p, NP-Num=s-ADJP-SubPOS=f, NP-Num=s-ADJP-Deg=p.

New features are established using constituency labels and morphological features of the phrase's head, as well as morphological features of the head and its dependent. To find the heads we only use the main POS tags in the case of the *Replace* method, these new features could only be applicable to *ExtendLex*. Here we used simple methods to find the heads of the nonterminals, but we could probably improve the results with more refind head finding rules.

4.4.2 Dependency-based Features

The SPMRL 2014 Shared Task had a dependency track. The organizers provided dependency annotations over the same texts. We used our team's dependency prediction [3] and created features from that. These features are made from heads of constituents and their dependency relations. We used features describing relations between the same head-dependent pairs in both the constituency and dependency parses. The features are described in detail in [11]. The frequency of these relations was also used.

These features are especially interesting for Hungarian because we have two manually annotated corpora in both representations as opposed to the other SPMRL languages [23].

4.4.3 Brown Cluster-based Features

We defined Brown cluster-based features. Brown clustering is a context-based hierarchical clustering over words [4]. Brown clusters are useful for syntax because words with similar context may have similar grammatical roles. These features can handle the feature sparsity issue. Utilizing these clusters, we duplicated every feature containing word by replacing each of those words with their Brown clusterID.

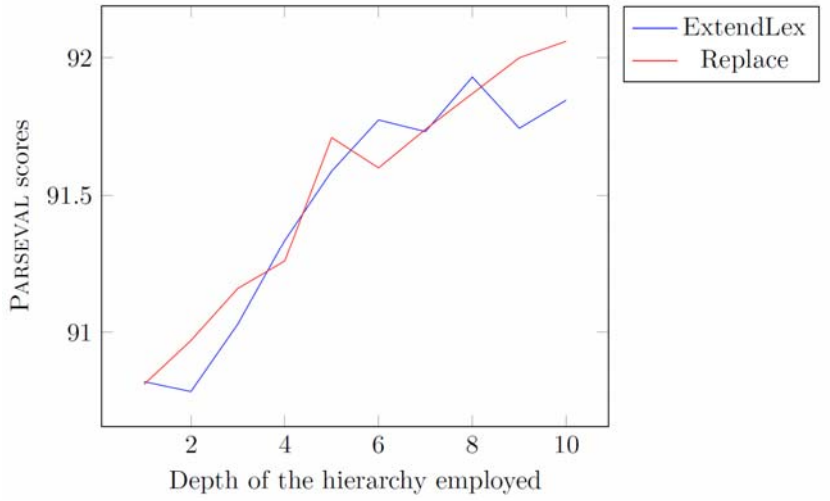

Figure 3
Results of Brown cluster based feature templates on the Hungarian dataset

Figure 3 investigates the effect of employing different levels of the Brown hierarchical tree evaluated on the Hungarian dataset. We obtained similar improvement in the case of both methods, *Replace* and *ExtendLex*, namely that these features increased the PARSEVAL metric with ~0.9 percentage point. We optimized the depth of hierarchy for each language separately.

5 Results and Discussion

As our baselines we used the results from Table 3. It contains the result of the Berkeley Parser with main POS and full morphological lexicon.

Table 4 shows the results achieved by the two strategies for handling lexical sparsity and the effect of the usage of the product of different grammars. The *Baseline* row contains the best results from Table 3. The product of the grammars increased the accuracy in every case. We obtained the best results with the

Replace Product system in six languages, but this strategy could not predict the full morphological description. In the case of French the *ExtendLex Product* obtained slightly better scores. In the next step we added the reranker stage to our product parsing systems.

Table 4
PARSEVAL scores on the development sets for the predicted setting

	baq	fra	ger	heb	hun	pol	swe
Baseline	77.82	79.35	82.26	88.71	87.18	86.75	75.19
ExtendLex	77.57	79.67	81.54	88.24	88.99	88.21	74.52
Replace	84.27	80.26	82.99	89.73	89.59	90.29	77.08
ExtendLex Product	79.47	**81.38**	82.94	89.22	90.43	91.52	78.21
Replace Product	**85.31**	81.29	**84.55**	**89.87**	**90.72**	**92.28**	**78.66**

Table 5 shows the final results of the reranker on the development set. We evaluated the effect of each new feature template and the combinations of all feature sets. *dflt* is the standard feature set from Charniak and Johnson [6] and Collins [10]. We added our morphology-based (*morph*), dependency-based (*dep*) and Brown cluster-based (*Brown*) features to this baseline. In the case of configurations, which contain Brown cluster based features, we show the best results in Table 5.

Table 5
PARSEVAL scores of the reranker on the development set for the predicted setting

ExtendLex	baq	fra	ger	heb	hun	pol	swe
Reranked *dflt*	79.16	81.92	83.01	89.39	91.06	87.89	79.09
Reranked *dflt+morph*	79.41	82.88	83.36	89.63	91.27	88.31	77.96
Reranked *dflt+dep*	81.89	82.65	84.83	90.28	91.88	91.84	79.43
Reranked *dflt+Brown*	80.63	82.49	84.33	90.30	91.93	89.37	78.42
Reranked *dflt+morph+dep+Brown*	82.69	82.62	85.16	90.64	92.05	91.10	79.57
Replace							
Reranked dflt	86.11	82.30	84.59	90.02	91.09	88.31	78.87
Reranked dflt+dep	86.73	82.78	86.05	90.47	91.89	90.53	79.38
Reranked dflt+brown	86.57	82.65	85.85	90.62	92.06	92.24	80.16
Reranked dflt+dep+brown	87.24	82.94	86.56	90.73	92.40	92.41	80.71

Reranking with default features improved the scores over product grammars both for *ExtendLex* and *Replace*. In the case of both representations, the combination of the proposed feature templates further increased our scores except for Swedish, where the morphology-based features and Brown cluster-based features also decreased the accuracy of the *ExtendLex* method, but when we added each new feature set, we obtained a slight improvement compared to the *dflt+dep* model. Finally, the *Replace* method obtained higher scores in every language. But in some languages (French, Hebrew, Hungarian) the difference is small and we can keep the morphological information with *ExtendLex*. We plan to combine these two fundamentally different methods in the future.

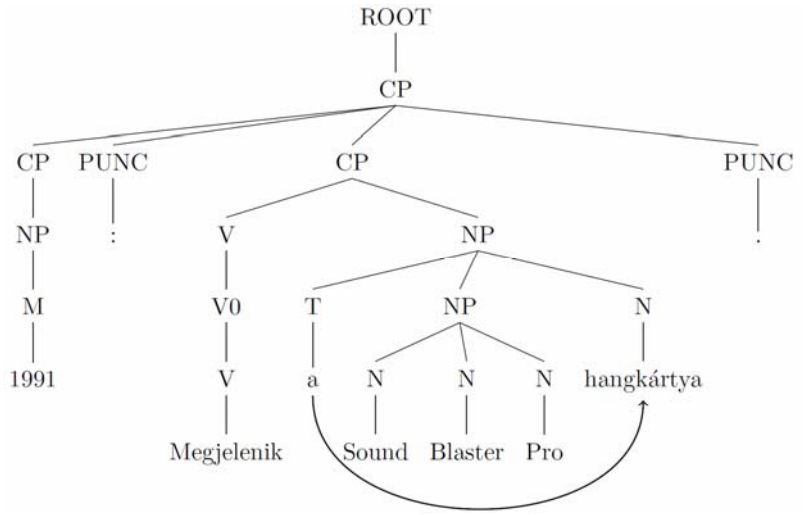

Figure 4

The gold standard parse of a Hungarian sentence with a dependency edge (*1991: The Sound Blaster Pro soundcard has appeared*)

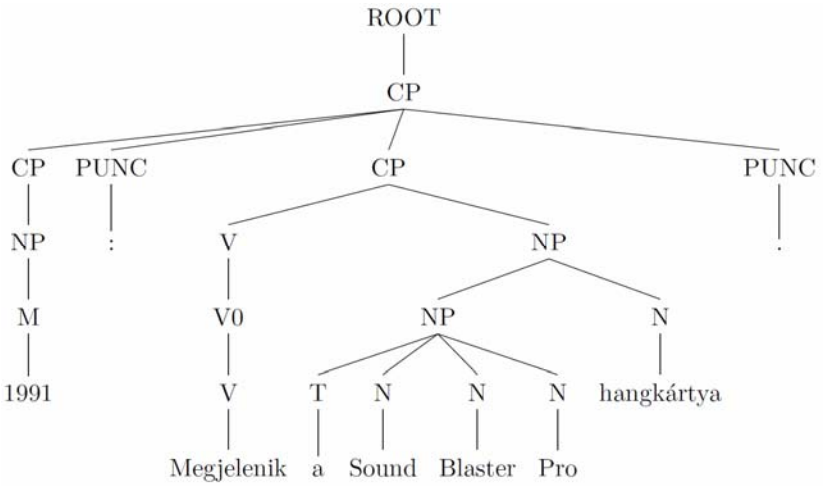

Figure 5

The reranked parse of a Hungarian sentence without dependency-based features

We manually analyzed the effect of new feature sets. Figures 4 and 5 show an example for the usefulness of the dependency based features. Figure 4 contains the correct parse of a Hungarian sentence. Figure 5 contains the result of our constituent parsing system to this sentence where we used the baseline (*dflt*) configuration. With dependency-based features (*dep*) we obtained the correct

parse. There is only one difference between the two parses. In the correct parse the article is connected to the noun phrase of "*Sound Blaster Pro hangkártya*" and in the wrong parse this article is connected to the noun phrase of "*Sound Blaster Pro*". If the computer does not see the dependency parse of this sentence, then the second parse is likely a good choice because a noun follows the article. But in the dependency parse there is a relation from the article to the noun *hangkártya*, and the dependency-based features encode this relationship, so the correct parse can be yielded.

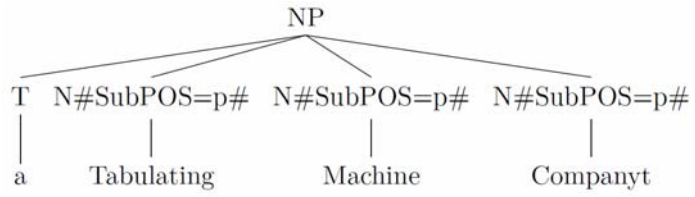

Figure 6
The gold standard parse of a part of a Hungarian sentence (the Tabulating Machine Company in accusative)

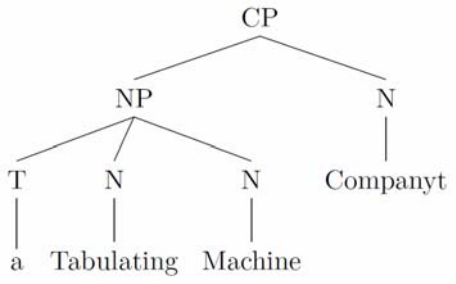

Figure 7
The reranked parse of a part of a Hungarian sentence without morphology-based features

Figure 6 and Figure 7 contain the parse of a part of a Hungarian sentence. In this phrase, "*Tabulating Machine Companyt*" is a proper noun, where "*Companyt*" is in the accusative case and the "a" is the article of the proper noun. In Hungarian a noun phrase can consist of the parts of a multi-word proper noun, like in Figure 6. Our parser (without morphology-based features) splits the expression because the parser has insufficient information to decide whether it is a multi-word proper noun, or it is just a sequence of independent nouns. In this expression, all wordforms are very rare. Instead of the usage of the wordforms, when we employ the morphology-based features, we can use binary features that encode whether the head of the NP (*Companyt*) is a proper noun and the head of the NP also has a sibling, which is a proper noun. Hence, these features can increase the probability of the correct sentence.

In the case of Brown-cluster based features we evaluated the F-score of the non-terminal labels. We found improvement in the case of labels ADJP (2.88), CP (1.39) and NP (0.86). The Brown clusters can group syntactically similar words, for instance, it clusterized the superlative adjectives and the ordinal numerals to one cluster. These words behave similarly at the syntax level since both groups usually function as the head of an ADJP. These automatically found similarities can help the reranker to choose the best parse from the candidates.

Conclusions

In this paper we introduced a system which achieved state-of-the-art results in constituent parsing of morphologically rich languages. We improved the efficiency of the discriminative reranking step with new feature templates. We proposed novel features from Brown clustering of the words and analysed the effect of morphology-based, dependency-based and Brown cluster-based features. The *Replace* method obtain better results for each of the languages, because this method can handle the lexical sparsity and also can handle the preterminal sparsity. The Brown cluster-based and the dependency-based features considerably contributed the system's performance and their beneficial effects overlap only to a small extent.

References

[1] 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. Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars. In E. Black, editor, Proceedings of the workshop on Speech and Natural Language, pp. 306-311, 1991

[2] Anders Björkelund, Özlem Çetinoğlu, Richárd Farkas, Thomas Müller and Wolfgang Seeker. (Re)ranking Meets Morphosyntax: State-of-the-art Results from the SPMRL 2013 Shared Task. In Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages, pp. 135-145, October 2013

[3] Anders Björkelund, Özlem Çetinoğlu, Agnieszka Faleńska, Richárd Farkas, Thomas Müller, Wolfgang Seeker, and Zsolt Szántó. Introducing the IMS-Wroclaw-Szeged-CIS entry at the SPMRL 2014 Shared Task: Reranking and Morpho-syntax meet Unlabeled Data. In Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages, pp. 97-102, August 2014

[4] Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jenifer C. Lai, and Robert L. Mercer. Class-based n-gram Models of Natural Language. Computational Linguistics, 18(4):467-479, 1992

[5] Eugene Charniak. A Maximum-Entropy-inspired Parser. In Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, pp. 132-139, 2000

[6] Eugene Charniak and Mark Johnson. Coarse-to-Fine n-Best Parsing and MaxEnt Discriminative Reranking. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05, pp. 173-180, 2005

[7] Xiao Chen and Chunyu Kit. Higher-Order Constituent Parsing and Parser Combination. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) pp. 1-5, 2012

[8] Key-Sun Choi, Young S Han, Young G Han, and Oh W Kwon. KAIST Tree Bank Project for Korean: Present and Future Development. In Proceedings of the International Workshop on Sharable Natural Language Resources, pp. 7-14, Citeseer, 1994

[9] Stephen Clark and James R. Curran. Wide-Coverage Efficient Statistical Parsing with CCG and Log-Linear Models. Computational Linguistics, 33, 2007

[10] Michael Collins. Discriminative Reranking for Natural Language Parsing. In Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00, pp. 175-182, 2000

[11] Richárd Farkas, Bernd Bohnet and Helmut Schmid. Features for phrase-structure reranking from dependency parses. In Proceedings of the 12th International Conference on Parsing Technologies, pp. 209-214, 2011

[12] Alexander Fraser, Helmut Schmid, Richárd Farkas, Renjing Wang and Hinrich Schütze. Knowledge Sources for Constituent Parsing of German, a Morphologically Rich and Less-Configurational Language. Comput. Linguist., 39(1):57-85, March 2013

[13] Yoav Goldberg and Michael Elhadad. Word Segmentation, Unknown-Word Resolution, and Morphological Agreement in a Hebrew Parsing System. Computational Linguistics, 39(1):121-160, 2013

[14] Liang Huang. Forest Reranking: Discriminative Parsing with Non-Local Features. In Proceedings of ACL-08: HLT, pp. 586-594, 2008

[15] Joseph Le Roux, Benoit Sagot and Djamé Seddah. 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, pp. 55-61, 2012

[16] Andrew Kachites McCallum. "Mallet: A Machine Learning for Language Toolkit". http://mallet.cs.umass.edu, 2002

[17] Thomas Mueller, Helmut Schmid and Hinrich Schütze. Efficient Higher-Order CRFs for Morphological Tagging. In Proceedings of the 2013

Conference on Empirical Methods in Natural Language Processing, pp. 322-332, 2013

[18] Slav Petrov. Coarse-to-Fine Natural Language Processing. PhD thesis, University of California at Bekeley, Berkeley, CA, USA, 2009

[19] Slav Petrov. Products of Random Latent Variable Grammars. In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 19-27, Association for Computational Linguistics, 2010

[20] Slav Petrov, Leon Barrett, Romain Thibaux and Dan Klein. 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, pp. 433-440, 2006

[21] Djamé Seddah, Sandra Kübler and Reut Tsarfaty. Introducing the SPMRL 2014 Shared Task on Parsing Morphologically-Rich Languages. In Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages, pp. 103-109, 2014

[22] Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Spence Green, Nizar Habash, Marco Kuhlmann, Wolfgang Maier, Joakim Nivre, Adam Przepiorkowski, Ryan Roth, Wolfgang Seeker, Yannick Versley, Veronika Vincze, Marcin Woliński and Alina Wróblewska. Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages. In Proceedings of the 4th Workshop on Statistical Parsing of Morphologically Rich Languages: Shared Task, Seattle, WA, 2013

[23] Katalin Ilona Simkó, Veronika Vincze, Zsolt Szántó and Richárd Farkas. An Empirical Evaluation of Automatic Conversion from Constituency to Dependency in Hungarian. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pp. 1392-1401, Dublin City University and Association for Computational Linguistics, 2014

[24] Zsolt Szántó and Richárd Farkas. Special Techniques for Constituent Parsing of Morphologically Rich Languages. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, pp. 135-144, Gothenburg, Sweden, April 2014