The greatest PoS tags, just the best really

Anonymous ACL submission

Abstract

Abstraction ensues.

1 Introduction

2 Previous/Related work

The lack of previous efforts on extrinsic evaluation of the effects of PoS tag sets on downstream applications is considerable. We will, instead/however, consider(/noe annet, considerable...consider) papers that evaluate the effects of PoS tag set granularity on PoS tagging. This includes investigating the effects of PoS tag sets on tagging of Swedish (Megyesi, 2001; Megyesi, 2002; Megyesi, 2009) and English (MacKinlay, 2005). Increases in tagger accuracy does not readily translate to improved parsing.

3 Data

3.1 The Norwegian Dependency Treebank

We used the newly developed Norwegian Dependency Treebank (NDT) (Solberg et al., 2014), which is the first publicly available treebank for Norwegian. It was developed at the National Library of Norway in collaboration with the University of Oslo, and contains manual syntactic and morphological annotation. The treebank contains 311 000 tokens of Bokml and 303 000 tokens of Nynorsk. The annotated texts are mostly newspaper text, but also include government reports, parliament transcripts and excerpts from blogs. The annotation process of the treebank was supported by the Oslo-Bergen Tagger and then manually corrected by annotators.

Noe om OBT? hmmm

Linguistic motivation, insight, utility...

Morphological Annotation The morphological annotation follows that of the Oslo-Bergen Tagger (Hagen et al., 2000; Solberg, 2013), which in turn is largely based on the work of Faarlund et al. (1997). The tag set of NDT is based on the tag set of the Oslo-Bergen Tagger. The PoS tags are complemented by a set of available morphological features, such as definiteness, number and tense. These features are used in our tag set modifications, where the coarse PoS tag of relevant tokens are concatenated with one or more of these features. This allows us to include more linguistic information in the tags. The tag set consists of 12 morphosyntactic PoS tags, with 7 additional tags for punctuation and symbols. The tag set is thus rather coarse-grained, with broad categories such as `subst` (noun) and `verb` (verb).

Head	**Dependent**
Preposition	Prepositional complement
Finite verb	Complementizer
First conjunct	Subsequent conjuncts
Finite auxiliary	Lexical/main verb
Noun	Determiner

Table 1: Annotation choices in NDT.

Syntactic Annotation The syntactic annotation choices in NDT are largely based on the Norwegian Reference Grammar (Faarlund et al., 1997). The annotation choices are outlined in Table 1, taken from Solberg et al. (2014), providing overview of the analyses of syntactic constructions that often distinguish dependency treebanks, such as coordination and the treatment of auxiliary and main verbs.

4 Experimental Setup

In preparation to conducting our experiments with linguistically motivated tag set modifications, a concrete setup for the experiments needed to be established, which is presented in the following.

Data Set Split As there was no standardized data set split of NDT due to its very recent development, we needed to establish a data set split (training/development/test) in preparation to our experiments. Our data set split of the treebank follows the standard 80-10-10 (training/development/test) split and will be distributed with the treebank and proposed as the new standard(?). In creating the data set split, care has been taken to preserve contiguous texts in the various data sets while keeping the split balanced in terms of genre (and source). Our proposed data set split was used in the Norwegian contribution to the Universal Dependencies project (vrelid and Hohle, 2016). The split will be made available at a companion website.

Tagger For our experiments with tag set modifications, we wanted a PoS tagger that is both fast and accurate. There is often a trade-off between the two, as the best taggers tend to suffer in terms of speed due to their complexity. However, a tagger that achieves both close to state-of-the-art accuracy as well as very high speed is TnT (Brants, 2000). The fact that TnT was used for evaluating the universal tag set (Petrov et al., 2012), served as another good indication of TnT being appropriate for our task. The sum of these factors led to TnT being the tagger of choice for our experiments.

Parser In choosing a syntactic parser for our experiments, we considered previous work on dependency parsing of Norwegian, specifically that of (Solberg et al., 2014). They found the Mate parser (Bohnet, 2010) to be the most successful parser for the parsing of NDT. Furthermore, recent dependency parser comparisons (Choi et al., 2015) showed that Mate performed very well on parsing of the English portion of the OntoNotes 5 corpus, beating a range of contemporary state-of-the-art parsers.

Tag Set Mapping In order to alter the tag set of NDT, we created a mapping for carrying out the tag set modifications. We created a mapping for carrying out the tag set modifications that maps the relevant existing tags to new, more fine-grained tags including more relevant morphological features for the applicable tokens.

Baseline It is common practice to compare the performance of PoS taggers to a pre-computed baseline for an initial point of comparison. For PoS tagging, a commonly used baseline is the Most Frequent Tag (MFT) baseline, which we use in our experiments. This involves labeling each word with the tag it was assigned most frequently in the training. All unknown words, i.e., words not seen in the training data, are assigned the tag most frequently assigned to words seen only once in the training. Unknown and infrequent words have in common that they rarely occur, and we might therefore expect them to have similar properties.

Tags & Features As we seek to quantify the effects of PoS tagging in a realistic setting, we want to run the parser on automatically assigned PoS tags. For the training of the parser, however, we have two options: using either gold standard or automatically assigned tags. In order to settle on a configuration, we conducted experiments with gold standard and automatically assigned tags to see how they differ with respect to performance. The results of our experiments reveal that the combination of training and testing on automatic tags is superior to training on gold standard tags and testing on automatic tags, surprisingly. Consequently, the parser was both trained and tested on automatically assigned tags in our experiments.

Note that it is absolutely crucial that the morphological features in the treebank are removed when using automatic tags, as they are still gold standard. For instance, if a verb token is erroneously tagged as a noun, we could potentially have a noun token with verbal features such as tense, which markedly obfuscates the training and parsing. Another important factor is that we want to isolate the effect of PoS tags, necessitating the exclusion of morphological features.

5 Tag Set Optimization

For each tag, we first experiment with each of the features in isolation before employing various combinations of them. We base our choices of combinations on how promising the features are and what we deem worth investigating in terms of linguistic utility, in order to see how the features might interact.

Category	Feature(s)	MFT	Accuracy	LAS	UAS
Baseline	—	94.14%	97.47%	87.01%	90.19%
Noun	Type, case & definiteness	89.61%	97.05%	88.81%	91.73%
Verb	Finiteness	93.72%	97.35%	87.30%	90.43%
Adjective	Degree	94.13%	97.41%	87.29%	90.44%
Determiner	Definiteness	94.13%	97.49%	87.30%	90.42%
Pronoun	Type & case	94.12%	97.51%	87.30%	90.41%

Table 2: Results of tagging and parsing with the most successful tag set modification for each category.

<table>
<tr><th>Tag</th><th>Description</th></tr>
<tr><td>adj|komp</td><td>Comparative adjective</td></tr>
<tr><td>adj|pos</td><td>Positive adjective</td></tr>
<tr><td>adj|sup</td><td>Superlative adjective</td></tr>
<tr><td>det|be</td><td>Definite determiner</td></tr>
<tr><td>det|ub</td><td>Indefinite determiner</td></tr>
<tr><td>pron|pers</td><td>Personal pronoun</td></tr>
<tr><td>pron|pers|akk</td><td>Personal pronoun, accusative</td></tr>
<tr><td>pron|pers|nom</td><td>Personal pronoun, nominative</td></tr>
<tr><td>pron|refl</td><td>Reflexive pronoun</td></tr>
<tr><td>pron|res</td><td>Reciprocal pronoun</td></tr>
<tr><td>pron|sp</td><td>Interrogative pronoun</td></tr>
<tr><td>subst|appell</td><td>Common noun</td></tr>
<tr><td>subst|appell|be</td><td>Common noun, definite</td></tr>
<tr><td>subst|appell|be|gen</td><td>Common noun, definite, genitive</td></tr>
<tr><td>subst|appell|ub</td><td>Common noun, indefinite</td></tr>
<tr><td>subst|appell|ub|gen</td><td>Common noun, indefinite, genitive</td></tr>
<tr><td>subst|prop</td><td>Proper noun</td></tr>
<tr><td>subst|prop|gen</td><td>Proper noun, genitive</td></tr>
<tr><td>verb|fin</td><td>Finite verb</td></tr>
<tr><td>verb|infin</td><td>Nonfinite verb</td></tr>
</table>

Table 3: The optimized tag set.

Tag set	MFT	Accuracy	LAS	UAS
Original	**94.14%**	**97.47%**	87.01%	90.19%
Full	85.12%	93.46%	87.13%	90.32%
Optimized	89.20%	96.85%	**88.87%**	**91.78%**

Table 4: Results of tagging and parsing with the optimized tag set, compared to the initial tag sets.

6 Optimized Pipeline

7 Summary/Conclusion and Future Work

References

Bernd Bohnet. 2010. Very High Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 89–97, Beijing, China.

Thorsten Brants. 2000. TnT - A Statistical Part-of-Speech Tagger. In *Proceedings of the Sixth Applied Natural Language Processing Conference*, Seattle, WA, USA.

Jinho D. Choi, Joel Tetreault, and Amanda Stent. 2015. It Depends: Dependency Parser Comparison Using A Web-Based Evaluation Tool. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 387–396, Beijing, China.

Jan Terje Faarlund, Svein Lie, and Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Universitetsforlaget, Oslo, Norway.

Kristin Hagen, Janne Bondi Johannessen, and Anders Nklestad. 2000. A Constraint-Based Tagger for Norwegian. In *Proceedings of the 17th Scandinavian Conference of Linguistics*, pages 31–48, Odense, Denmark.

Andrew MacKinlay. 2005. The Effects of Part-of-Speech Tagsets on Tagger Performance. Bachelor's thesis, University of Melbourne, Melbourne, Australia.

Beta Megyesi. 2001. Comparing Data-Driven Learning Algorithms for PoS Tagging of Swedish. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, pages 151–158, Pittsburgh, PA, USA.

Beta Megyesi. 2002. *Data-Driven Syntactic Analysis: Methods and Applications for Swedish*. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden.

Beta Megyesi. 2009. The Open Source Tagger HunPoS for Swedish. In *Proceedings of the 17th Nordic Conference on Computational Linguistics*, pages 239–241, Odense, Denmark.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A Universal Part-of-Speech Tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 2089–2096, Istanbul, Turkey.

Per Erik Solberg, Arne Skjrholt, Lilja vrelid, Kristin Hagen, and Janne Bondi Johannessen. 2014. The Norwegian Dependency Treebank. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 789–795, Reykjavik, Iceland.

Per Erik Solberg. 2013. Building Gold-Standard Treebanks for Norwegian. In *Proceedings of the 19th Nordic Conference of Computational Linguistics*, pages 459–464, Oslo, Norway.

Lilja vrelid and Petter Hohle. 2016. Universal Dependencies for Norwegian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, Portorož, Slovenia.