Morphosyntactic Tagging of Slovene: Evaluating Taggers and Tagsets

Sašo Džeroski,* Tomaž Erjavec,* Jakub Zavrel†

* Dept. for Intelligent Systems, Jožef Stefan Institute
Ljubljana, Slovenia
{saso.dzeroski,tomaz.erjavec}@ijs.si
† Centrum voor Nederlandse Taal en Spraak, University of Antwerp
Antwerp, Belgium
jakub.zavrel@kub.nl

Abstract

The paper evaluates tagging techniques on a corpus of Slovene, where we are faced with a large number of possible word-class tags and only a small (hand-tagged) dataset. We report on training and testing of four different taggers on the Slovene MULTEXT-East corpus containing about 100.000 words and 1000 different morphosyntactic tags. Results show, first of all, that training times of the Maximum Entropy Tagger and the Rule Based Tagger are unacceptably long, while they are negligible for the Memory Based Taggers and the TnT tri-gram tagger. Results on a random split show that tagging accuracy varies between 86% and 89% overall, between 92% and 95% on known words and between 54% and 55% on unknown words. Best results are obtained by TnT. The paper also investigates performance in relation to our EAGLES-based morphosyntactic tagset. Here we compare the per-feature accuracy on the full tagset, and accuracies on these features when training on a reduced tagset. Results show that PoS accuracy is quite high, while accuracy on Case is lowest. Tagset reduction helps improve accuracy, but less than might be expected.

1. Introduction

Trainable wordclass syntactic taggers have reached the level of maturity where many models and implementations exist, with several being robust and available free of charge. In this context, comparing, evaluating, and tuning taggers for 'new' languages becomes imperative. Recently, there has been growing interest in validation of language resources and processing tools, some explicitly concerned with tagger evaluation, e.g. the GRACE project (Adda et al., 1998) for French.

Less work on tagger evaluation and tagging in general has been done on the so-called Eastern European languages. These languages typically have quite different properties, in particular much richer word inflection. Standard morphosyntactic tagsets are therefore orders of magnitude larger (from 600 to over 3000); an even greater problem is the lack of training and testing data, i.e., pre-annotated corpora. The acquisition of training data, of course, gets easier when at least basic automatic methods are in place.

The situation is beginning to change, in part due to the results of the MULTEXT-East project (Dimitrova et al., 1998) which, for Czech, Romanian, Hungarian, Estonian, Bulgarian and Slovene developed common language resources. These contain EAGLES-based morphosyntactic descriptions (Erjavec and Monachini, eds., 1997), medium sized word-form lexica utilising these descriptions (Ide et al., 1998) and a small parallel corpus annotated with disambiguated lexical information (Erjavec and Ide, 1998). These hand-validated resources have been used as a starting point in a number of experiments on tagging and tagset design, e.g., for Romanian (Tufiş, 1999; Tufiş, 2000) and Hungarian (Varadi, 1999). Tagging methods have also been developed and tested for Czech (Hajič and Hladka, 1998a; Hajič and Hladka, 1998b); recently, an evaluation of tagging was performed on the complete multilingual MULTEXT-East updated resources (Hajič, 2000), including Slovene.

In our work on tagging evaluation (Džeroski et al., 1999) we concentrate on the Slovene portion of the MULTEXT-East corpus, on which we trained and tested four different taggers. The taggers tested were rule-based tagger (Brill, 1995), the Maximum Entropy Tagger (Ratnaparkhi, 1996), the Memory-based Tagger (Daelemans et al., 1996) and the tri-gram tagger TnT (Brants, 1999).

The motivation for the work comes from the need for having an evaluation of tagger performance on Slovene language data, and for obtaining baseline results with 'standard' taggers. Further methods can then be used to boost performance (tagger or tagset combinators), or the accuracy numbers can be used as a benchmark against which to compare newly developed taggers. While we have not performed tests on other MULTEXT-East languages, we believe that results would carry over at least to the more similar of the languages, e.g., Czech. This is confirmed by comparable error rates on Slovene reported by us and by (Hajič, 2000).

We also investigate tagging performance on Slovene in relation to our EAGLES-based morphosyntactic tagset. Here we compare the per-feature accuracy on the full MSD tagset, and experiment with tagset reductions, a practice commonly adopted to improve tagger performance.

Section 2 presents the MULTEXT-East Slovene resources used in the experiment, and Section 3 the tagger evaluation experiment and the synopsis and analysis of testing results. Section 4 deals with minimising errors by modifying (reducing) the tagset. Section 5 gives conclusions.

2. The Data

The Slovene MULTEXT-East resource most relevant to tagging is the translation of Orwell's '1984'. The corpus is tokenised, and its words marked for morphosyntactic descriptions (MSDs) and lemmas. For our dataset we took the revised version of the resources published on the CD-ROM (Erjavec et al., 1998). This section presents the Slovene morphosyntactic descriptions and the annotated corpus that served as the datasets used for training and testing the taggers and tagsets.

2.1. Morphosyntactic descriptions

The syntax and semantics of the MULTEXT-East MSDs are given in the *morphosyntactic specifications* of the project (Erjavec and Monachini, eds., 1997). These specifications have been developed in the formalism and on the basis of specifications for six Western European languages of the EU MULTEXT project (Bel et al., 1995). These common specifications were developed in cooperation with EAGLES, (Calzolari and McNaught, eds., 1996).

The MULTEXT-East morphosyntactic specifications contain, along with introductory matter also:

1. the list of defined categories (parts-of-speech)
2. common tables of allowed attribute-value pairs
3. language specific tables

Of the MULTEXT-East *categories*, Slovene uses Noun (N), Verb (V), Adjective (A), Pronoun (P), Adverb (R), Adposition (S),[1] Conjunction (C), Numeral (M), Interjection (I), Abbreviation (Y), Particle (Q) and Residual (X).[2] .

The *common tables* give, for each category, a table defining the attributes appropriate for the category, and the values defined for these attributes. They also define which attributes/values are appropriate for each of the MULTEXT-East languages; the tabular structure facilitates the addition of new languages. The format of the common tables is exemplified by the start of the Noun table, given in Table 1.

The common tables have a strictly defined format, which enables the automatic expansion and validation of MSDs. For example, according to the tables, the MSD `Pg-nsg----n` is valid for Slovene and expands to `Pronoun general neuter singular genitive nominal`

The *language specific tables* are, again, organised by category, and provide commentary on the attributes and values for a particular language, as well as feature co-occurrence restrictions and exhaustive lists of valid MSDs. The Slovene tables also contain localisation information, which enables automatic translation of the MSDs into Slovene: the above `Pg-nsg----n` translates to `Zc-ser----s / Zaimek celostni srednji ednina rodilnik samostalniški`.

```
Noun (N)

11 Positions

**** **** **** **** ****
PoS  Type Gend Numb Case
**** **** **** **** ****
---- ---- ---- ---- ---- ----
Def  Cltc Anim OwnN OwnP OwdN
---- ---- ---- ---- ---- ----

= ====== ============ =  EN RO SL CS BG ET HU
P ATT    VAL          C  x  x  x  x  x  x  x
= ====== ============ =
1 Type   common       c  x  x  x  x  x  x  x
         proper       p  x  x  x  x  x  x  x
- ------ -----------  -
2 Gender masculine    m  x  x  x  x  x
         feminine     f  x  x  x  x  x
         neuter       n  x  x  x  x  x
- ------ -----------  -
3 Number singular     s  x  x  x  x  x  x  x
         plural       p  x  x  x  x  x  x  x
         dual         d        x  x
         count        t              x
- ------ -----------  -
4 Case   nominative   n        x  x  x  x  x
         genitive     g        x  x     x  x
         dative       d        x  x        x
         accusative   a        x  x        x
         vocative     v     x     x  x
         locative     l        x  x
         instrumental i        x  x        x
         direct       r     x
         oblique      o     x
         partitive    1                 x
         illative     x                 x  x
         inessive     2                 x  x
         elative      e                 x  x
...
```

Table 1: Example of MSD Table: Nouns

For an impression of the information distribution of the Slovene MSDs we give in Table 2 for each category, four values: the number of appropriate *attributes* of the category; the total number of *values* for all its attributes; the number of different MSDs in annotated Slovene '1984' corpus; and the number of different MSDS in the *lexicon*, which contains the full inflectional paradigms for all its lemmas.

The table shows that almost half MSDs that are in the lexicon do not appear in the corpus; this reflects the small size of the corpus, but also the grammar-like orientation of the MSDs. With pronouns, for example, it is often the case that a certain MSD describes only a single lexical entry, and an infrequent one at that.

2.2. The Dataset

For our dataset we took the Slovene '1984' corpus, in particular the first three parts of the novel; we held back the Appendix. The corpus is pre-segmented and pre-tokenised, and each word is annotated with its context-disambiguated MSD; punctuation is tagged as well, starting with X. We

[1] Adpositions include prepositions and postpositions; Slovene uses only prepositions.

[2] Residual is a MSD category encompassing unknown (unanalysable) lexical items and is not used for words in the corpus. In our experiments we used it to mark punctuation symbols.

PoS	Att	Val	1984	Lexicon
Pronoun	11	36	594	1,335
Adjective	7	22	169	279
Numeral	7	23	80	226
Verb	8	26	93	128
Noun	5	16	74	99
Preposition	3	8	6	6
Adverb	2	4	3	3
Conjunction	2	4	2	3
Interjection	0	0	1	1
Abbreviation	0	0	1	1
Particle	0	0	1	1
Σ	45	139	1,025	2,083
Punctuation	1	10	10	-

Table 2: Slovene morphosyntactic distribution

split the corpus into ten random folds, held back fold 0, then used fold 1 for testing, and folds 2–9 for training.[3]

As shown in Table 3, the dataset has about 6000 sentences and 100.000 tokens, and is split into 90% training and 10% testing data.

	Full	Train	Test
Sentences	5855	5204	651
Tokens	92399	81805	10594
Words	77772	68825	8947
Ambigs	87.2%	86.4%	70.2%
Diff pairs	18649	17166	3912
Diff words	16017	14831	3573
Diff MSDs	1004	976	543

Table 3: Corpus dataset

Tokens are either punctuation or words, and the former comprise cca 15% of the tokens. Of the word tokens in the corpus, around 80% are MSD ambiguous. The final three rows give the numbers of different word/MSD pairs, of words only, and of MSDs.

As compared to the training set, the testing set contains 1245 (11.75%) previously unseen tokens. Furthermore, 300 tokens in the testing have been seen in the training set, but never with the MSD assigned to them in the testing set.

3. Comparing Taggers

Different taggers were tested on the dataset, making use of a simple regime of training and testing: the complete context dependent as well as context independent (lexical) knowledge about Slovene came from the training corpus. Each tagger was trained on this data, with their various parameters left at their default values. The taggers were tested on the test set tokens (words and punctuation), and the accuracy computed. The experiment thus makes no use of a background lexicon: when testing accuracy on unknown words and annotations, unknown-ness is determined w.r.t. the lexicon derived from the training set.

[3]The dataset used is available from the Slovene 'Learning Language in Logic' site, *http://nl.ijs.si/lll/*

3.1. The taggers

The four taggers tested on our dataset represent different popular tagging approaches. The choice was made on the basis of their availability to the authors. Furthermore, they had to satisfy the following conditions:

- ability to handle large tagset
- ability to tag unknown words
- ability to induce lexicon from training data
- being robust

3.1.1. RBT

The Rule Based Tagger was written by Eric Brill of John Hopkins University (Brill, 1992; Brill, 1994; Brill, 1995). The tagger starts with a base annotation of the corpus, and searches for a sequence of transformation rules that "repair" errors. The base annotation is to assign each word its most frequent tag. Unknown words are initialised as nouns; the tagger first learns a set of rules for unknown words and then a set of contextual rules for all words. Rules are generated until they correct no more than a certain number n of errors. This threshold is hard-coded in the source code.

3.1.2. MET

The Maximum Entropy Tagger, written by Adwait Ratnaparkhi, builds a probabilistic model from the family of exponential models: $p(tag_i, context) = \pi\mu \prod_j \alpha_j^{f_j(tag_i,context)}$ where f_j are binary features defined on the combination of a tag and some (simple or complex) property of the context. The templates from which features are generated are described in (Ratnaparkhi, 1996). During the training of the tagger, a numeric optimisation method called Improved Iterative Scaling (IIS) is used to find the weights α_j for the features. Features which occur less than ten times are not considered. The default number of iterations of IIS is 100. Once the tagger is trained it does an n-best search for the best tag sequence.

3.1.3. MBT

This tagger, written by Jakub Zavrel, Peter Berck and Walter Daelemans of Tilburg University, is described in detail in (Daelemans et al., 1996). The MBT tagger stores examples (cases) from the corpus in memory and constructs a classifier (IGTREE) that assigns tags to new text by extrapolation from the most similar examples in memory. First a lexicon is constructed from the corpus, and this lexicon is converted into ambiguity classes (multi-tags). An ambiguity class is an ordered set of tags that a word can take, where tags that fall below a certain threshold (e.g. 10 %) are omitted. A separate classifier is constructed for known and unknown words. The cases for known words have the features ddfWaa, which means two disambiguated tags to the left of the focus word, the ambiguity class of the focus word, the focus word itself (only the 100 most frequent words are included in this feature) and two ambiguous tags to the left. The cases for the unknown words have the features chndFasss, meaning has-capital, has-hyphen, has-number, one disambiguated tag to the left of the focus word, one ambiguous tag to the right and three suffix letters. The classifier for the unknown words is constructed only from words that have a frequency lower than 5 in the training corpus.

3.1.4. TnT

TnT, short for Trigrams'n'Tags (Brants, 1999), is a very efficient statistical part-of-speech tagger that is trainable on different languages and virtually any tagset. The component for parameter generation trains on tagged corpora. The system incorporates several methods of smoothing and of handling unknown words. TnT is not optimised for a specific but rather for training on a large variety of corpora. Adapting the tagger to a new language, new domain, or new tagset is very easy. Additionally, TnT is optimised for speed.

The tagger is an implementation of the Viterbi algorithm for second order Markov models. The main paradigm used for smoothing is linear interpolation, the respective weights are determined by deleted interpolation. Unknown words are handled by a suffix trie and successive abstraction.

3.2. Error analysis

The accuracies were computed using a Black-Box Combiner (van Halteren et al., 1998; Džeroski et al., 1999) that trains and tests all specified taggers on the same dataset, flags the results and computes the accuracies on all words, and separately on known tokens (word/tag and punctuation/tag pairs seen in the training corpus), on unknown words, and on words that are known, but for which the correct MSD had not been seen in the training set. Table 4 gives a synopsis of the results for the four taggers.

Type of test	RBT	MET	MBT	TnT
Known, OK	8405	8285	8468	8604
Known, err	644	764	581	445
Unk. word, OK	701	472	687	848
Unk. word, err	544	773	558	397
Unk. MSD, OK	0	91	0	0
Unk. MSD, err	300	209	300	300

Table 4: Tagging results

The accuracies in per-cent are given in Table 5; there unknown words are taken to be both the those that have not been seen, as well as those that only had a new MSD in the testing data. For the 300 such cases, it is only MET that resolves about a third of them correctly; the other taggers all treat the induced ambiguity class of a words as complete, a mistake that increases their overall error rate more than a third. To overcome this limitation, a background morphological lexicon covering all possible MSDs of the words in the training corpus would be needed.

Apart from accuracy, the question of training and testing speed is also paramount; here RBT was by far the slowest (over a day for training), followed by MET, with MBT and TnT being very fast (both less than 1 minute). Language models are easier to tune with fast taggers, and this can then also lead to increased accuracy.

For each tagger the accuracy was tested not only on the MSDs, but also on isolated features of the feature-structure-like MSDs. Full MSDs are used for learning and are also predicted. These predictions are projected on the isolated features to obtain the feature predictions. Table 5 gives, for all, known and unknown tokens accuracies on full MSDs, and also on part-of-speech. Additionally we give for known words accuracies on the Type, Case, Number and Gender attributes. The accuracies were computed only for tokens for which the relevant feature was in fact appropriate.

Token type	Tokens	RBT	MET	MBT	TnT
All	10594	85.95	86.36	86.42	89.22
on PoS	10594	95.64	94.66	95.31	96.59
Known	9049	92.88	91.56	93.58	95.08
on PoS	9049	98.75	97.02	98.76	98.51
on Type	8713	98.67	96.94	98.82	98.71
on Case	3557	87.74	88.16	88.89	93.06
on Number	4629	97.19	96.28	97.43	98.33
on Gender	4556	95.90	93.99	96.62	97.65
Unknown	1545	45.37	55.92	44.47	54.88
on PoS	1545	77.41	80.84	75.08	85.30

Table 5: Tagging accuracies

The table shous that PoS accuracy, esp. for known words is quite high and comparable to that achieved by taggers e.g., for English. This is of course due to the small tagset, but also to the relativelly low PoS ambiguity of Slovene words. Conversely, words are much more inflectionally ambiguous, and these features, esp. Case, are much harder to predict.

Of course, the accuracies are quite different depending on whether the token is a punctuation symbol (X) or a verb, noun or adjective. Especially interesting are part-of-speech accuracies for unknown words; using a background lexicon we can cover pronouns and numerals of Slovene, but no lexicon can cover productive words, i.e., verbs, and especially nouns and adjectives. Table 6 gives the number of tokens and accuracy attained by the TnT tagger on all the tokens in the testing set, and split into known and unknown.

PoS	All tokens		Known		Unknown	
	n	%	n	%	n	%
Σ	10594	89.2	9049	95.0	1545	54.8
X	1647	100.0	1647	100.0	0	-
V	2454	95.8	2044	99.0	410	79.7
N	1901	81.4	1356	92.9	545	53.0
P	1062	79.0	1014	82.7	48	0.0
C	828	96.4	828	96.4	0	-
S	811	96.1	807	96.6	4	0.0
A	757	61.6	316	90.8	441	40.8
R	696	93.9	629	96.3	67	71.6
Q	336	88.6	332	89.7	4	0.0
M	98	65.3	72	83.3	26	15.3
I	3	100.0	3	100.0	0	-
Y	1	100.0	1	100.0	0	-

Table 6: PoS Tagging Accuracies

The Table shows that Verb accuracy is in fact quite good, but Noun and esp. Adjective are below average.

4. Tagset reductions

We also conducted some experiments on tagset design, where we decreased the cardinality of the tagset by either omitting certain attributes, or omitting almost all, except certain attributes. The rationale behind this is that it might be easier to predict smaller (less complex) tags than highly complex ones. The MBT tagger was used to perform 9-fold cross-validation on folds 1-9 mentioned earlier. Table 7 lists the tagsets considered, their cardinality, and the accuracies of MBT (averaged over the 9 folds). The accuracies are on all (known and unknown) tokens.

Tagset	Cardinality	MBT Accuracy
PoS Only	12	96.07
Type Only	38	95.57
All but Case	392	89.67
All but Gend	582	88.22
All but Numb	602	86.94
All but Type	665	87.27
Full MSDs	1021	86.93

Table 7: MBT accuracies on reduced tagsets

The accuracies are in a similar order to tagset cardinality; the less tags, the better the results. However, the accuracy gain is less than might be expected: tagging with the full MSD set and predicting PoS only gets 95.31%, while tagging with PoS only 96.07%, a relative gain of only 16%. Obviously, richer tags also give a richer context for correct disambiguation. In line with per-feature accuracy, it is also the tagset with Case omitted that shows the best performance. Taking a closer look at full MSD predictions of MBT on known tokens projected on the Case attribute shows that the Case of prepositions is the easiest to predict (accuracy of 93.74%), and the case of numerals hardest (accuracy of 73.61%).

We also trained the Combiner (van Halteren et al., 1998) for which the attributes are the tags predicted by MBT for each tagset and the class the correct maximal tag. The testing results of MBT on partitions 1-9 were used for training. The obtained combiner was tested on partition 0, combining the MBT predictions for the individual tagsets into a single prediction (in the maximal tagset). The results were only slightly better than the ones obtained by using the maximal tagset tagger alone.

5. Conclusions

The article presented experiments on applying machine learning based tagging approaches to the MULTEXT-East Slovene corpus. These initial results indicate that the trigram-based TnT tagger is probably the best choice considering both accuracy (especially on unknown words) and efficiency, followed by the memory based MBT tagger. Using more resource intensive tagging approaches as RBT or MET does not bring accuracy advantages, at least with our large MSD tagset and using their default features.

The MSD tagset can be reduced to increase performance, although not proportionally the diminishing number of tags. Selective feature removal from the MSDs shows that inflectional features are much harder to predict than lexeme ones, and that the Case attribute is the most difficult to determine.

The results obtained provide a baseline to which more sophisticated approaches should be compared.

Acknowledgements

The work presented here was supported by the ESPRIT IV project 20237 **ilp2** and by INCO/COPERNICUS projects COP-106 MULTEXT-East and PL96-1142 CONCEDE.

6. References

Adda, G., J. Mariani, J. Lecomte, P.Paroubek, and M. Rajman, 1998. The GRACE French Part-Of-Speech Tagging Evaluation Task. In *First International Conference on Language Resources and Evaluation, LREC'98*. Granada: ELRA.

Bel, N., N. Calzolari, and M. Monachini (eds.), 1995. Common specifications and notation for lexicon encoding and preliminary proposal for the tagsets. MULTEXT Deliverable D1.6.1B, ILC, Pisa.

Brants, Thorsten, 1999. Tnt – statistical part-of-speech tagging. http://www.coli.uni-sb.de/˜thorsten/tnt/

Brill, Eric, 1992. A simple rule-based part of speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing, ACL*. Trento, Italy.

Brill, Eric, 1994. Some advances in transformation-based part-of-speech tagging. In *Proc. of AAAI'94.*.

Brill, Eric, 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–565.

Calzolari, N. and J. McNaught (eds.), 1996. Synopsis and Comparison of Morphosyntactic Phenomena Encoded in Lexicons and Corpora: A Common Proposal and Applications to European Languages. EAGLES Document EAG—CLWG—MORPHSYN/R, ILC, Pisa. http://www.ilc.pi.cnr.it/EAGLES/home.html.

Cutting, D., J. Kupiec, J. Pedersen, and P. Sibun, 1992. A practical part-of-speech tagger. In *Proceedings of the Third Conference on Applied Natural Language Processing*. Trento, Italy.

Daelemans, W., J. Zavrel, P. Berck, and S. Gillis, 1996. MBT: A memory-based part of speech tagger-generator. In Eva Ejerhed and Ido Dagan (eds.), *Proceedings of the Fourth Workshop on Very Large Corpora*. Copenhagen.

Dimitrova, L., T. Erjavec, N. Ide, H.-J. Kaalep, V. Petkevič, and D. Tufiş, 1998. Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages. In *COLING-ACL '98*. Montréal, Québec, Canada.

Džeroski, S., T. Erjavec, and J. Zavrel, 1999. Morphosyntactic tagging of slovene: Evaluating pos taggers and tagsets. Research Report IJS-DP 8018, Jožef Stefan Institute, Ljubljana. http://nl.ijs.si/lll/bib/dzerza-report/

Erjavec, T. and M. Monachini (eds.), 1997. Specifications and notation for lexicon encoding. MULTEXT-East Final Report D1.1F, Jožef Stefan Institute, Ljubljana. http://nl.ijs.si/ME/CD/docs/mte-d11f/

Erjavec, T. and N. Ide, 1998. The MULTEXT-East corpus. In *First International Conference on Language Resources and Evaluation, LREC'98*. Granada: ELRA.

Erjavec, T., A. Lawson, and L. Romary, 1998. East meets West: A Compendium of Multilingual Resources. CD-ROM. ISBN: 3-922641-46-6.

Hajič, Jan, 2000. Morphological Tagging: Data vs. Dictionaries. In *ANLP/NAACL 2000*. Seatle.

Hajič, Jan and Barbora Hladka, 1998a. Czech Language Processing / POS Tagging. In *First International Conference on Language Resources and Evaluation, LREC'98*. Granada: ELRA.

Hajič, Jan and Barbora Hladka, 1998b. Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structured Tagset. In *COLING-ACL'98*. Montréal.

Ide, Nancy, Dan Tufiş, and Tomaž Erjavec, 1998. Development and Assessment of Common Lexical Specifications for Six Central and Eastern European Languages. In *First International Conference on Language Resources and Evaluation, LREC'98*. Granada: ELRA.

Ratnaparkhi, Adwait, 1996. A maximum entropy part of speech tagger. In *Proc. ACL-SIGDAT Conference on Empirical Methods in Natural Language Processing*. Philadelphia.

Tufiş, Dan, 1999. Tiered Tagging and Combined Language Model Classifiers. In Jelinek and Noth (eds.), *Text, Speech and Dialogue*, number 1692 in Lecture Notes in Artificial Intelligence. Springer.

Tufiş, Dan, 2000. Using a Large Set of EAGLES-compliant Morpho-Syntactic Descriptors as Tags for Probabilistic Tagging. In *Second International Conference on Language Resources and Evaluation, LREC'00*. Athens: ELRA.

van Halteren, H., J. Zavrel, and W. Daelemans, 1998. Improving data driven wordclass tagging by system combination. In *COLING-ACL'98*. Montréal.

Varadi, Tamas, 1999. Morpho-syntactic ambiguity and tagset design for hungarian. In *Proceedings of the EACL-99 Workshop on Linguistically Interpreted Corpora (LINC-99)*. Bergen: ACL.