How to Decrease Performance of a Statistical Parser
On negative influence of some outside factors

Daniel Zeman

Abstract

This paper is a study of outside factors that may have a negative impact on a parser's accuracy. The discussed phenomena can be divided into two classes, those related to treebank design, and those related to morphological tagging. Although the scope of the paper is limited to the Prague Dependency Treebank, two particular taggers and one particular parser, we believe that our observations may be of interest to future treebank designers.

Anyway, since training parsers is usually not the only purpose of building treebanks, other good reasons may prevail and push aside the parser's preferences. For that case, we suggest how the parser may use nasty input. More fine-grained elaboration of these workarounds and their evaluation is a matter of future work.

1. Introduction

Statistical parsers heavily depend on treebanks. Large amounts of data are needed for training (not to speak of testing) a parser, and non-statistical parsers can profit from treebanks as well. Unfortunately, building a treebank is simply a too much complicated and expensive effort to tailor it just to one purpose — and for most treebanks it certainly holds that supporting parsers was *not the only* reason why they appeared. While preferences sometimes clash, a consensus has to be found and the result may not always be the optimal fulfilment of a parser designer's dreams.

It is apparent that not knowing the background of negotiating the guidelines for a particular treebank, including all the pros and contras, one can hardly criticize the designers from just one perspective — and I'm by no means attempting that. But as statistical parser cannot be built until a version of the treebank exists, it is difficult to simulate and defend its needs during the design phase of the treebank building. One purpose of this paper is to remind future treebank designers of some classes of parsing-complicating features they should avoid if possible. Vice-versa, the purpose also is to reminder parser designers of the traps a treebank may have prepared for them. And finally, I want to present some ideas of how these traps might be dealt with, although a deeper elaboration and evaluation of these workarounds is a matter of future work.

The second part of the paper will discuss another important factor that has an impact on parsing accuracy: the quality of morphological disambiguation, performed by taggers. Partially this matter is still related to the treebank (or tagset and dictionary, in this case) design, but mostly it is just limited by the state of the art of tagging.

The investigated language is Czech. While we hope that our ideas are portable to other related languages, language-dependent phenomena are still to be expected. All of the observations and experiments we describe were conducted on the Prague Dependency Treebank 1.0 (Böhmová et al. 2000, Hajič et al. 1999). The parser accuracy was measured (where not stated otherwise) on PDT 1.0 analytical development test data (7319 non-empty sentences, 126030 words). Observations that did not employ any tool trained on PDT were done on PDT 1.0 analytical training data (73088 non-empty sentences, 1255590 words). PDT 1.0 provides morphological disambiguation by two taggers, identified as "a" and "b" (tagger "a": a maximum-entropy-based tagger, see Hajič and Hladká 1998; tagger "b": an HMM-based tagger described in Hajič et al. 2001 (but without the rule-based module described there)). We will discuss the contribution of both taggers later in the paper.

The parser used in all experiments is the one of Zeman (2002).

2. Problem one: compound prepositions

The usual parsing of prepositional phrases in PDT is designed, according to Hajič et al. (1999), as follows: the preposition is the head of the phrase, and the noun phrase depends on it. E.g., *přechod z obrany do útoku,* "transition from defense to attack", is parsed as *přechod (z (obrany) , do (útoku))*:
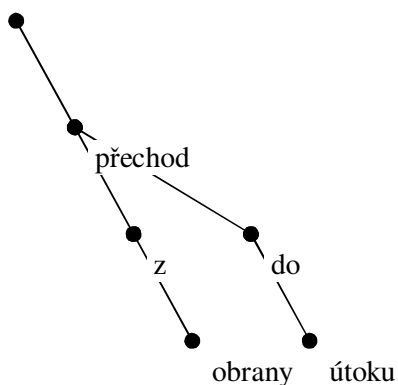

Figure 1: přechod z obrany do útoku

Hajič et al. (1999) also introduce a notion of *improper* or *secondary preposition*. It is a word originally belonging to a different part of speech, which functions as a preposition in some sentences. For instance, the word *počátkem* is normally a noun in instrumental case (*počátek* "beginning"). Some of its instances are really regarded as such: *Počátkem krize se nezabýval, protože na něm nemohl nic změnit.* "He was not interested in the beginning of the crisis as he was not able to change anything on it." But in other examples *počátkem* is syntactically a preposition, although the morphological analyzer does not think so: *Počátkem března začala krize.* "The crisis began in the beginning of March." This case still does not do any serious harm to the parser since many nouns take other nouns in genitive as complements and the parser can handle *počátkem* as just another one of those.

The bad news is that a secondary preposition can also be formed by a *sequence* of two or three words, at least one of which usually is a primary preposition. For some reason, the guidelines demand
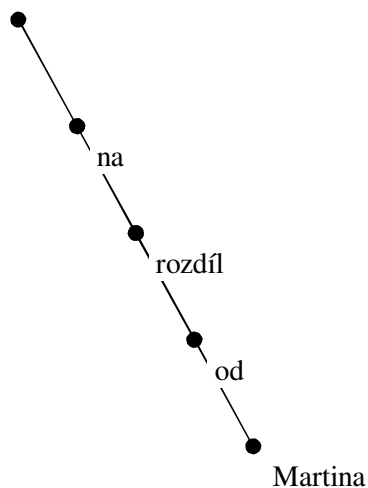

Figure 3: na rozdíl od Martina (correct)
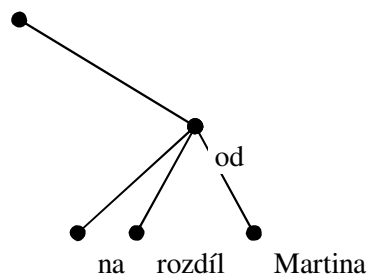
Figure 2: na rozdíl od Martina (intuitive)

the annotators to parse such sequences in a counter-intuitive manner, as the primary preposition hangs as a leaf on what would otherwise be its child. For instance, there is the secondary preposition *na rozdíl od* "in contrast to", composed by the (primary) preposition *na* "on/in", its noun *rozdíl* "contrast", and another primary preposition *od* "from" (here "to") that connects the compound preposition with the following noun phrase. The intuitive and the correct parses of *na rozdíl od Martina* "in contrast to Martin" are shown in Figure 2 and 3, respectively.

Of course, any statistical model would prefer the intuitive parse, as it is an analogy to all normal prepositional phrases.

Unfortunately we cannot just take the list of secondary prepositions and tell the parser to use the strange structure whenever such a word sequence appears. They can also occur in other functions and be parsed normally[1]! We don't think it is a good approach to change structure on the level of surface syntax just because something has a different (deep) function. There are primary prepositions that bear different functions as well[2], and the distinction is made only on the tectogrammatical level.

Statistics: To find out how much harm this inconsistency really does or does not to the parser, we collected some statistics. There are 126030 words in the test data, out of which 12380 are prepositions by morphology (m-tag starting with R). Should the preposition be defined as having the s-tag AuxP (prepositional function on analytical level; in our example, all words except of *Martin* would get that tag), there were 12558 such prepositions. More precisely, 176 morphologically defined prepositions (m-preps) appeared as non-prepositions on the analytical level, and 354 words tagged with AuxP (s-preps) were not m-preps at the same time. Both cases may have occasionally been caused by an error in automatic m-tagging (the s-tagging in all our data has been done manually).

The overall accuracy of the parser (measured as the number of correctly assigned governing nodes divided by the total number of words) is 71.1 %. The accuracy of finding governors for m-preps is 64.5 % (63.9 % for s-preps, 64.8 % for ms-preps (both m- and s-preps at the same time), and 42.6 % for m!-preps (m-preps not marked as s-preps at the same time)). The accuracy of hanging words that *should have hung* on an m-prep is 91.0 % (89.7 % for s-preps).

The guidelines list 150 secondary prepositions, out of which 22 are made of three words. All the three-word ones are of the RNR type (m-prep – noun – m-prep; examples: *bez ohledu na* "without respect to", *na rozdíl od* "in contrast to", *v souvislosti s* "in connection with", *ve vztahu k* "in relation to"). There further are 70 two-word ones, most of them (60) of the RN type (*k rukám* "to the hands of", *po dobu* "for the time of", *s výjimkou* "with the exception of", *z hlediska* "from the point of view of" etc.). Other types include VR (verb – m-prep: *soudě podle* "judging according to") or DR (adverb – m-prep: *společně s* "together with"). The remaining 58 secondary prepositions are one-word and thus not critical for the structural analysis. 27 of them even have the option of being m-prep, although the tagger usually considers other options, too, and may not be able to pick the right one at all times. The others are most frequently nouns, sometimes also adverbs or transgressives.

Fortunately for the parser, only some of the listed word sequences are found repeatedly in PDT, forming quite a small fraction of its volume. The most represented type is RN as there are 171 occurrences (342 words) in test data (and another 39 not listed in the guidelines but structured the same way, either because of their foreign origin (*de facto, in Prague, of America, van Miert*; see the section 4.1 of this paper) or because of a misclassification based on the m-tags (*Pro Thalia, Via Dolorosa, o Dá*). There were 35 varieties of that type, the most frequent examples being *v rámci* "in the framework of" (29), *v případě* "in case of" (22), and *na základě* "on the basis of" (14). The RNR type occurred 84

---

[1] *Petr na rozdíl od Pavla přišel.* "Petr came, in contrast to Pavel." (parsed as secondary preposition) vs. *Petr na rozdíl mezi Janou a Jitkou nehledí.* "Petr does not mind the difference between Jana and Jitka." (parsed as primary preposition)
[2] *Synové zdědili* **po** *miliónu.* "The sons inherited a million each." vs. *Synové zdědili milióny* **po** *otci.* "The sons inherited millions from the father."

times, in 13 varieties, and the most frequent examples were *na rozdíl od* "in contrast to" (21), and *ve srovnání s* "in comparison with" (11).

Evaluation of the RNR type: 84 times the PDT requested that an m-prep depend on the second word right (another m-prep). The parser fulfilled that request at 19.0 %. The word between the two m-preps shall also depend on the second m-prep. This request was fulfilled at a quite surprising level, 64.3 %.

Evaluation of the RN type: 210 times the PDT requested that an m-prep depend on its right neighbor. The parser fulfilled that request at 18.6 %. Vice versa, the noun is now the happy head of the phrase, and can depend on virtually everything but not its fellow preposition, what is what the parser assumes in most cases. So, the accuracy on such dependencies is as low as 7.6 %, while it would be some 78.1 % if the guidelines for prepositions were consistent. (If we throw away the misclassifications as *de facto* and test also the nounness of the second word (which discards the RR type *vzhledem k*), 154 occurrences remain and the contrast gets even stronger: correctly hung are 0.6 % of nouns, making the guidelines consistent would result in 98.7 %.

Of course, the small representation of compound prepositions in the data means that even such alarming difference as just mentioned has only little impact on the overall accuracy. However, if the theory is consistent, there will be less noise in the training data and further parsing errors will be repaired. So in our last experiment we "amended" each training RNR or RN instance to look like as if no notion of secondary prepositions existed. We did that before feeding the training data into the parser, and so we did to the test data before using it for testing. The resulting overall accuracy was 71.4 %. Compared to the baseline of 71.1 %, 456 dependencies were corrected.

Conclusion: A treebank should give the compound prepositions the same structural annotation as is found with normal prepositions. If this is not possible, the parser may use the list of secondary prepositions, but it has to check that the compound preposition does not occur at the end of the sentence and that the noun in the compound preposition is not modified. Even then some errors may arise, and the statistical nature of the parser will be hurt.

3. Problem two: Numerals

The numerals in Czech are a challenge for linguists. The main obstacle with them is that they sometimes agree with counted nouns in case, sometimes they don't. Following this inconsistency in language, the PDT designers defined two completely different structural annotations for counted phrases: one headed by the number, and another headed by the counted noun. Obviously this is much more of challenge for a parser, as it has to capture the agreement rule from the data. It's not easy in most cases because the case of words is often tagged incorrectly and the statistics are heavily biased. And should the amount be expressed by a number instead of a numeral word, it is practically impossible to pick the correct structure, since numbers have no endings serving as case clues and the case is not annotated for numbers at all (of course, if an automatic tagger attempted to tag it, it would be highly error-prone).

Let's have a closer look at the various structural annotations now. The numerals *jeden* "one", *dva* "two", *tři* "three", and *čtyři* "four" always agree with their nouns in case. Thus *3 koblihy* "3 donuts" will be parsed: donuts (3).

The numerals *pět* "five" and more do agree when the whole phrase is in genitive, dative, locative or instrumental but will not agree if it's nominative, accusative or vocative. Should one of these three cases
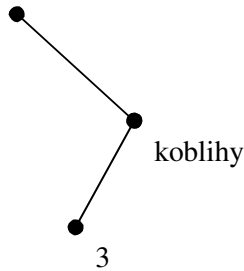

Figure 4: 3 koblihy

be applied to the whole phrase, the numeral will take the form common to all these three cases, while the counted phrase will obligatorily take the genitive form. Thus if the numeral is replaced by a number (no case marking at all) and the counted phrase is in genitive, we have absolutely no clue whether the whole phrase is in genitive as well (and headed by the noun) or it is in nominative or accusative (and headed by the numeral). The only help might result from looking at the governor of the numeral phrase (provided the parser is able to pick it correctly at the moment). If it subcategorizes for genitive (e.g. some prepositions and verbs do), the phrase can be marked as genitive. Unfortunately, too much good fortune is needed to achieve all that automatically, and the question is: Was the distinction between the agreeing and non-agreeing cases so important to deserve such effort? We believe it wasn't.
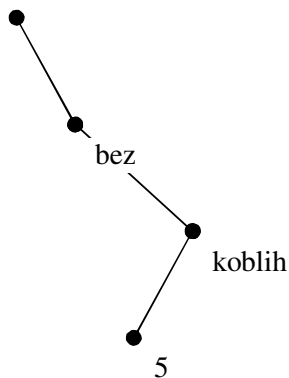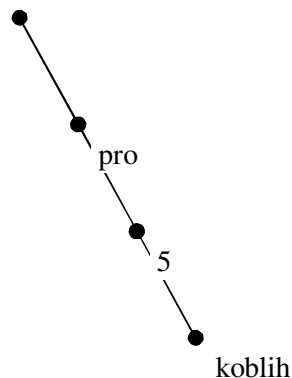

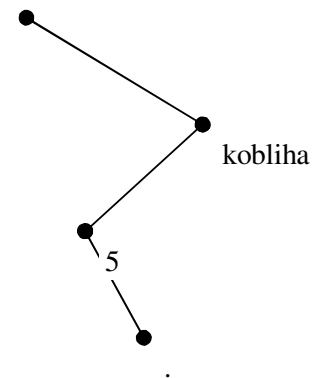*Figure 5: bez 5 koblih*　　*Figure 6: pro 5 koblih*　　*Figure 7: 5 . kobliha*

One more remark: if the number is followed by a dot, it may (but need not) represent an ordinal numeral. Those are similar to adjectives and always agree with the ranked nouns in case. Thus an ordinal number never governs a noun.

Example: *pro 5 koblih* "for 5 donuts" will be parsed "for 5 (donuts)" but *bez 5 koblih* "without 5 donuts" will be parsed "without (donuts (5))". *5. kobliha* "5th donut" is parsed as "donut (5 (.))".

Statistics: There are 3435 occurrences of numerals (morphological tag beginning with C) in PDT test data (126030 words). Out of these, 2030 (59.1 %) do not show case, and 1887 (54.9 %) are

expressed using Arabic numbers. 263 occurrences belong to the numbers 1, 2, 3, and 4; the rest (1624 occurrences) belong to numbers 5 and more, 0, and decimal numbers. Among the numerals expressed by a word, 811 appeared in nominative or accusative (as judged by the "a" tagger), 299 in genitive, and 294 in dative, locative or instrumental.

Evaluation: The parser achieves 71.0 % on hanging numerals, which is fairly close to its overall accuracy. In hanging words that *should have hung* on numerals, it achieves 66.4 %.

The structure of numeral phrases could be made consistent in either of two ways. Either the counted noun should always be head, or the number should.

Problem: If we require that any numeral be dependent on the counted noun we say by that that a non-genitive phrase can have got a genitive head, filling a non-genitive slot in something's subcategorization frame. This is obviously the reason why the treebank has numerals sometimes up, sometimes down. We believe however that the parser will be able to overcome this problem easier than the present one, just by remembering the case information from the numeral with the head. Such a case transfer has to be done elsewhere as well (in coordination headed by the conjunction, for instance).

We don't have an evaluation of the case transfer improvement but the results of parsing without it show that it will be needed. If all numerals are annotated as dependent on the counted things, the parser accuracy slightly drops to 70.9 %. However, the accuracy of annotating numerals rose to 92.2 %. If all numerals are annotated as governing the counted things, the overall accuracy is 70.1 %.

Conclusion: the numeral should be always dependent, which will at least guarantee consistency inside of the numeral phrase. Again, if it is not possible to influence the treebank design, the parser might rearrange the training trees so that they conform to the above condition, train on them, parse the test sentences, and as a final step, rearrange the output trees. At that moment it will at least be able to see whether a genitive-subcategorized preposition governs the phrase.

4. Other problems

There are other problems we know about but haven't done such an analysis as for the preceding two. Nevertheless, it is worth mentioning the remaining here.

4.1. Foreign words

The annotation guidelines say that phrases in a foreign language have to be annotated uniformly, all words as attributes of the last one. Of course, one cannot require that the annotators know virtually
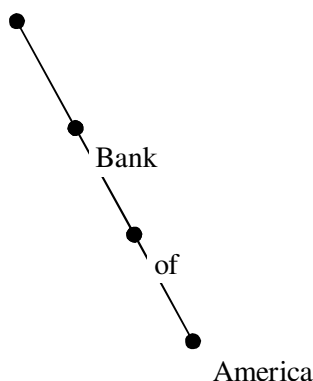

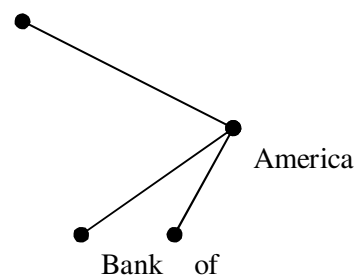*Figure 8: Bank of America (intuitive)*

Figure 9: Bank of America (correct)

every language to be able to describe the inner structure of such phrases. However, there is absolutely no clue for the parser to know whether a word is foreign or Czech. If it is not found in the dictionary,[3] it might be a Czech word not covered by the dictionary, although such cases are rare. If it is found in the dictionary, there is no way of identifying it as foreign. Thus *Bank of America* "Bank of America" is morphologically annotated as NNIS4 RR--X NNFS1 (noun preposition noun).[4] The parser naturally treats it as any other instance of such pattern, hanging *America* under *of*. However, this lowers its success rate because according to the foreign phrase rule *Bank* and *of* shall depend on *America*.

4.2. Nominal predicate vs. passive mood

The main auxiliary verb in Czech is *být* "to be"; whenever used as auxiliary, it is placed in the tree as a leaf under the meaning verb it modifies. Many complex verb forms are built using forms of *být*, including: *budu dělat* "I will do", *dělal bych* "I would do", *dělal jsem* "I did", *bylo uděláno* "it has been
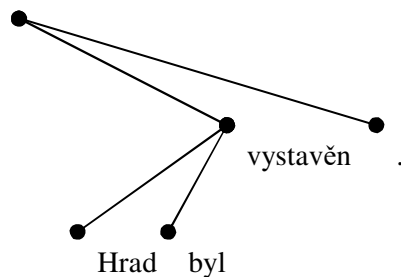

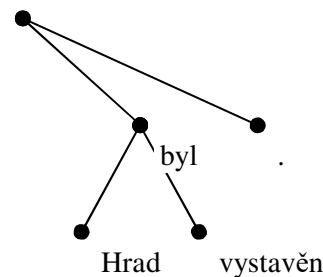Figure 10: Hrad byl vystavěn . (pass)   Figure 11: Hrad byl vystavěn . (pnom)

done". The last one — passive — is hardly distinguishable from a copula with a nominal predicate. Hajič et al. (1999), part "Distinguishing state and passive", say that passive mood (action) requires *být* to depend on the lexical verb; while nominal predicate (state) requires the lexical verb to depend on *být*. Thus two equal phrases can be parsed in two contrasting ways, and even a human judge needs a high level of semantic expertise — including the knowledge of the sentences in the neighborhood — to choose the right one.

Example: *Hrad byl vystavěn.* "The castle was built up."

4.3. *atd.* "etc." in coordinations

The chapter about coordinations in the annotation guidelines of PDT (Hajič et al. 1999) says: "It may happen that the conjunction *a (and)* is a part of some abbreviation (*apod.* and similarly, *atd.* etc.). Since the abbreviation is written without spaces and as such is represented by a single node, the function Coord is assumed by the whole abbreviation." In fact, there are annotations following that definition. Unfortunately, there are numerous examples of a contradicting approach as well and I'm not able to figure out why. In the first of the following examples, *atd.* is the head of the coordination, complying with the guidelines. In the second example, it violates the cited guideline.[5]

---

[3] We mean by that: it has not been assigned a meaningful morphological tag, i.e. it has a tag starting with X.

[4] One may question how does it come that the English words are found in a Czech dictionary. As I'm not the author of the dictionary, I can only tell for the word *bank*: there is such Czech word, meaning "jackpot" in English, not "bank". The other two (*of, America*) are not Czech — "America" is spelled *Amerika* in Czech. Probably their frequency in Czech newswire text is high enough to include them in the dictionary. It would be reasonable to mark them as foreign but unfortunately there is currently no such marking.

[5] The contradicting example was found using the on-line tree viewer Netgraph (see the link to PDT) by searching for [form=atd]. At least first three trees found violated the guideline (cb25am.fs#7, cb31am.fs#51, cc04am.fs#27).

Example 1: *energetická zařízení, stavební stroje, ruční nářadí atd.* "power plant devices, construction machines, hand tools etc."
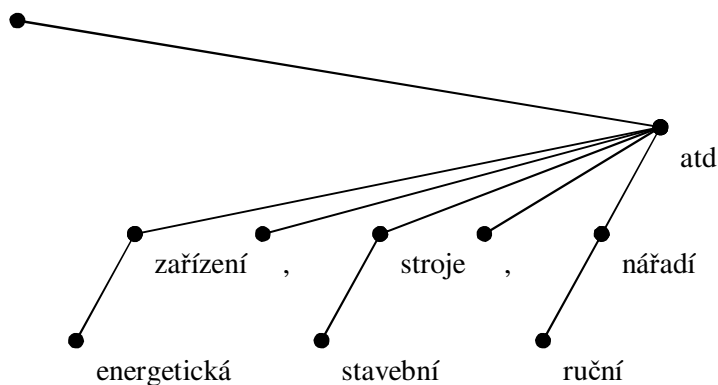

Figure 12: energetická zařízení , stavební stroje , ruční nářadí atd

Example 2: *konkurenci, technických novinkách v oboru, cenách srovnatelného zboží atd.* "competition, new technology in the field, prices of comparable products etc."
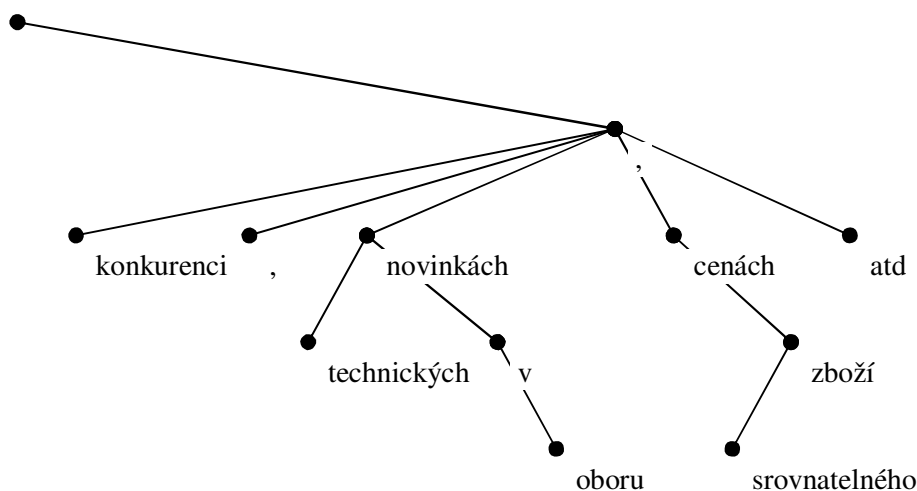

Figure 13: konkurenci , technických novinkách v oboru , cenách srovnatelného zboží atd

5. Tagging errors

The accuracy of a parser that relies on morphological tags essentially depends on the success rate of the tagger. Any tagging error may violate agreement in gender, number or case, which often are the determining factors for syntactic relations. The case tagging errors are the most crucial.

Hajič and Hladká (1998) published the error rate of the maximum entropy tagger (tagger "a"): 6.2 %, i.e. the accuracy is 93.8 %. They also measured the error rate of separate morphological attributes of each word. For us what is important is the accuracy of predicting the attributes used in our reduced tag set: case (95.2 %) and subpart of speech (99.5 %). We counted the same on our training data.[6] We got

---

[6] We could not include our test data in the experiment because it does not contain manually annotated morphology.

92.6 %[7] overall for tagger "a", and 92.7 % for tagger "b". Both numbers are better when our tag reducing scheme is applied (see Hajič et al. 1998): 94.5 % "a", 94.4 % "b". On the other hand, assigning the correct case is one of the more difficult tasks of tagging. We tested only the words whose correct case was known (i.e. 1-7, not X nor -): 91.7 % "a", 91.5 % "b". From the point of view of the parser it is also interesting in how big part of *all* words we can expect a case tagging error. The correct cases divided by all words give 95.3 % "a" and 95.2 "b".

Example: *Podnikatelskou misi do Kolumbie, Peru, Argentiny a Paraguaye připravila na dny 25. října až 8. listopadu hospodářská komora ČR.* "The businessmen's mission to Colombia, Peru, Argentina, and Paraguay, was prepared for the days 25th October till 8th November by the Chamber of Commerce of the Czech Republic." The four names of South-American countries are coordinated and agree in genitive case required by the preposition *do* "to". The tagger was confused and assigned genitive, unknown, genitive, and nominative, respectively. No wonder that the parser was confused even more and constructed a useless tree structure. The Figure 14 shows the tree output from the parser, the
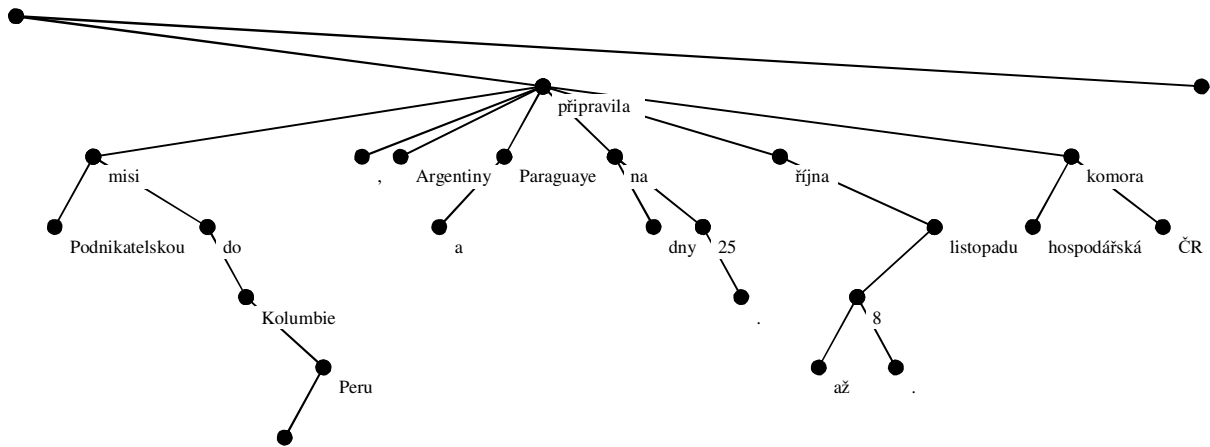

Figure 14: Podnikatelskou misi do Kolumbie , Peru , Argentiny a Paraguaye… (parser output)

Figure 15 shows the correct tree.

One of the possibilities of working around the tagging errors is to use other sources of morphological information. There are four different sources in PDT: the morphological analyzer (but its output is not disambiguated), the taggers "a" and "b", and the manual annotation. The taggers have just been discussed, and the manual annotation will not be available in the parsing phase, so the ambiguous morphology is the main hope. For the sake of completeness, we have tested parsing accuracy on the manual annotation as well.

Hajič et al. (1998) compared the parsing accuracy of an earlier version of our parser while gradually using ambiguous morphology, hand-annotated morphology, and output of a tagger. Various combinations of different sources for training and testing were tested as well. The worst combination proved to be ambiguous training – ambiguous parsing (51.4 %). The best one was tagger training – tagger parsing (54.1 %). Since then, the amount of available data has risen several times, the parser has improved significantly, and the tag set is reduced more drastically for the current parser, so we felt it was worth repeating that test.

---

[7] The tagging accuracy is measured on all words, not only the ambiguous ones but also words where the tagger had nothing to solve. Such accuracy is important from the point of view of a parser, which needs to know how good is its input. If we tested only the ambiguous words, we could say that the tagger was successful at 88.0 %. However, we still would not be able to distinguish between the words where the tagger had to choose one of two possibilities and the words where there were two dozens of choices.
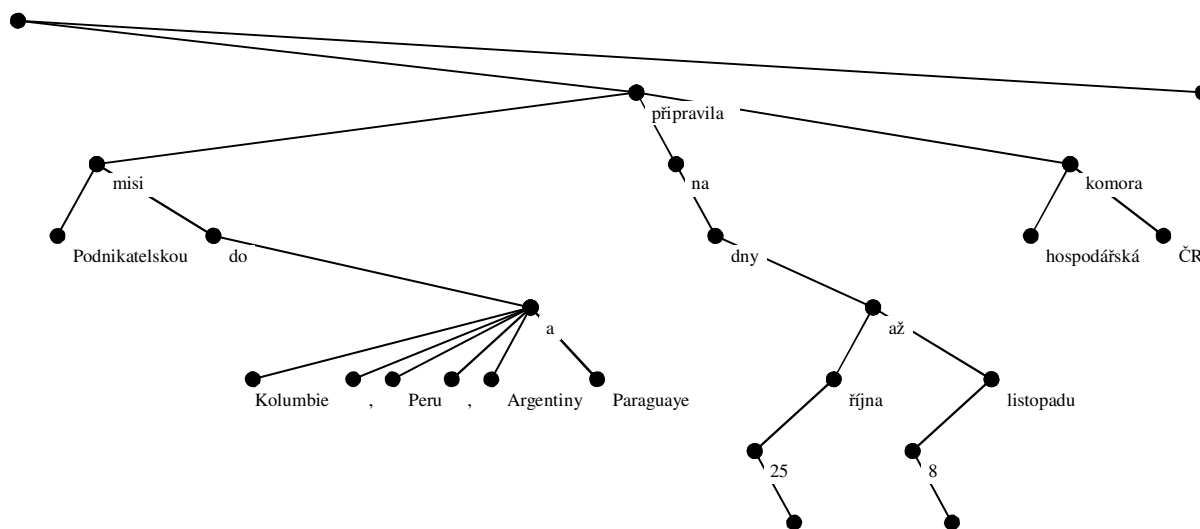
Figure 15: Podnikatelskou misi do Kolumbie , Peru , Argentiny a Paraguaye... (correct parse)

We did not use the development test data of the PDT. Instead, we split the training data so that we put aside every tenth sentence and kept the rest. This way we obtained new test data that contained hand annotated tags. The new data are also more representative as to the sources of the texts.[8] The new training data contains 65847 sentences and 1133509 words (tokens). The new test data contains 7241 sentences and 122081 words.

Evaluation: Using the morphology of tagger "a", the parser achieved an accuracy of 70.4 %. Using the manually annotated morphology, the accuracy rose to 71.9 % (that result is not of much use because the manual annotation will not be available for new data, but it is interesting as the previous test in Hajič et al. (1998) showed that manual – manual morphology gave worse results than tagger – tagger).

Finally, the ambiguous morphology can be used in a couple of ways. The simplest one (not tested in Hajič et al. 1998) is to look at the sequence of possible morphological tags as one string, one long tag. Of course, duplicates removed and list of tags ordered. When trained and tested on such multi-tags, the parser performed at 67.6 %. Another option is the one tested in Hajič et al. (1998). We counted 1/n of an occurrence of each tag X_i whenever a word occurred that could have been tagged by one of n tags $X_1...X_n$. Sum of relative frequencies of tags was used during parsing. That approach led to the surprisingly high accuracy of 71.8 % (only 1 ‰ less than on manual tags, and 1.4 % better than on tagger output!).

Last but not least, we could enrich this approach with our own partial tagging procedure. Whenever the parser selects a dependency the morphological situation of the words involved gets a bit clearer. For instance, if the dependency observes agreement in case, and the words allow for the tag combinations N1|N4|N5 and N1|N2|N4, the pairs N1–N1 and N4–N4 will contribute with a much higher probability amount than the other combinations. Thus the probability of the words having one of the tags N1, N4 (as opposed to N2 and N5) will get higher. Such information could than be used in finding the other dependencies. This method has not yet been tested but we will test it in the near future as we expect interesting results.

Conclusion: We have shown that even a small tagging error can badly affect parsing. However, the experiments promise that without any tagger we can do at least as well as with it.

---

[8] PDT contains texts from 4 sources. One of them, *Vesmír*, has been known to be much more difficult to parse than the 3 others (see Hajič et al. 1998). However, the PDT 1.0 development test data does not contain texts from this source, unlike the training data.

6. Acknowledgements

References

Alena Böhmová, Jan Hajič, Eva Hajičová, Barbora Hladká (2000) *The Prague Dependency Treebank: Three Level Annotation Scenario.* In: Anne Abeillé (ed.): Treebanks: Building and Using Syntactically Annotated Corpora. At: http://shadow.ms.mff.cuni.cz/pdt/. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Jan Hajič, Eric Brill, Michael Collins, Barbora Hladká, Douglas Jones, Cynthia Kuo, Lance Ramshaw, Oren Schwartz, Christoph Tillmann, Daniel Zeman (1998) *Core Natural Language Processing Technology Applicable to Multiple Languages. The Workshop 98 Final Report.* At: http://www.clsp.jhu.edu/ws98/projects/nlp/report/. Center for Language and Speech Processing, Johns Hopkins University, Baltimore, Maryland.

Jan Hajič, Barbora Hladká (1998) *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structured Tagset.* In: Proceedings of the 36th Annual Meeting of the ACL and the 17th International Conference on Computational Linguistics (COLING-ACL 98), vol. 1, pp. 483 – 490. Université de Montréal, Montréal, Québec.

Jan Hajič, Jarmila Panevová, Eva Buráňová, Zdeňka Urešová, Alla Bémová (1999) *Annotations at Analytical Level* (instructions for annotators, English translation by Zdeněk Kirschner). At: http://shadow.ms.mff.cuni.cz/pdt/Corpora/PDT_1.0/Doc/aman-en/. Univerzita Karlova, Praha, Czechia.

Jan Hajič, Pavel Krbec, Pavel Květoň, Karel Oliva, Vladimír Petkevič (2001) *Serial Combination of Rules and Statistics: A Case Study in Czech Tagging.* In: Proceedings of the 39th Annual Meeting of the ACL (ACL-EACL 2001). Université de Sciences Sociales, Toulouse, France.

Daniel Zeman (2002) *Can Subcategorization Help a Statistical Dependency Parser?* In: Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002), vol. 2, pp. 1156 – 1162. Zhongyang yanjiuyuan 中央研究院 (Academia Sinica), Taibei 台北, Taiwan.