

Czech-English Phrase-Based Machine Translation

Ondřej Bojar¹, Evgeny Matusov², and Hermann Ney²

¹ Institute of Formal and Applied Linguistics*
ÚFAL MFF UK, Malostranské náměstí 25, CZ-11800 Praha, Czech Republic
`bojar@ufal.mff.cuni.cz`

² Lehrstuhl für Informatik 6, Computer Science Department
RWTH Aachen University, D-52056 Aachen, Germany
`{matusov, ney}@cs.rwth-aachen.de`

Abstract. We describe experiments with Czech-to-English phrase-based machine translation. Several techniques for improving translation quality (in terms of well-established measure BLEU) are evaluated. In total, we are able to achieve BLEU of 0.36 to 0.41 on the examined corpus of Wall Street Journal texts, outperforming all other systems evaluated on this language pair.

1 Introduction

We aim at Czech-to-English machine translation (MT). For the time being, top performing systems of machine translation are statistical and phrase-based.¹

Czech is a thoroughly studied Slavonic language with extensive language data resources available (most notably the Prague Dependency Treebank, PDT², [1]). Czech is an inflective language with rich morphology and relatively free word order allowing non-projective constructions. These properties usually cast some doubt on the applicability of “uninformed” statistical methods that do not attempt at analyzing sentence structure.

Traditionally, most of the research on Czech is performed within the framework of the Functional Generative Description (FGD, [2]), a dependency-based formalism defining the deep syntactic (syntactico-semantic) level of language description. Effort has been invested in the development of linguistically adequate annotated data (PDT and lexicons) and tools (taggers, parsers to surface and deep syntactic levels, see the PDT for references). MT is attempted at the deep syntactic level [3].

In this paper, we describe our experiments with a phrase-based statistical MT system (PBT) developed at RWTH Aachen University [4]. We observe that at least for our particular corpus, translation direction and metrics used, linguistically uninformed methods currently clearly outperform other approaches.

* The work was performed while the first author was a visiting scientist at RWTH Aachen University.

¹ <http://www.nist.gov/speech/tests/summaries/2005/mt05.htm>

² <http://ufal.mff.cuni.cz/pdt2.0/>

1.1 Statistical Phrase-Based Machine Translation (Summary)

In statistical MT, the goal is to translate a source (foreign) language sentence $f_1^J = f_1 \dots f_j \dots f_J$ into a target language (English) sentence $e_1^I = e_1 \dots e_j \dots e_I$. Among all possible target language sentences, we choose the sentence with the highest probability:

$$\hat{e}_1^I = \operatorname{argmax}_{I, e_1^I} \{Pr(e_1^I | f_1^J)\} \quad (1)$$

In a log-linear model, the conditional probability of e_1^I being the translation of f_1^J is modelled as a combination of independent feature functions $h_1(\cdot, \cdot) \dots h_M(\cdot, \cdot)$ describing the relation of the source and target sentences:

$$Pr(e_1^I | f_1^J) = \frac{\exp(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J))}{\sum_{e_1^{I'}} \exp(\sum_{m=1}^M \lambda_m h_m(e_1^{I'}, f_1^J))} \quad (2)$$

The model scaling factors λ_1^M are trained either to the maximum entropy principle or optimized with respect to the final translation quality measure.

Among feature functions used, the most important are the phrase-based translation model and the target language model. The phrase-based model captures the basic idea of phrase-based translation: to segment source sentence into phrases, then translate each phrase and finally compose the target sentence from phrase translations. Theoretically, the segmentation s_1^K of the source sentence into K phrases is introduced as a hidden variable to the overall model (thus making the feature functions dependent also on the segmentations, i.e. $h(f_1^J, e_1^I, s_1^K)$) and summing over all possible segmentations. In practice, a maximum approximation to this sum is used:

$$h_{\text{Phr}}(f_1^J, e_1^I) = \max_{s_1^K} \log \prod_{k=1}^K p(\tilde{f}_k | \tilde{e}_k) \quad (3)$$

The conditional probability of phrase \tilde{f}_k given phrase \tilde{e}_k is estimated from relative frequencies: $p(\tilde{f}_k | \tilde{e}_k) = N(\tilde{f}, \tilde{e}) / N(\tilde{e})$ where $N(\tilde{f}, \tilde{e})$ denotes the number of co-occurrences of a phrase pair (\tilde{f}, \tilde{e}) that are consistent with the word alignment. The marginal count $N(\tilde{e})$ is the number of occurrences of the target phrase \tilde{e} in the training corpus.

The phrase-based model is included in the log-linear combination in source-to-target and target-to-source directions: $p(\tilde{f} | \tilde{e})$ and $p(\tilde{e} | \tilde{f})$. In addition, statistical single word based lexica are used in both directions. They are included to smooth the relative frequencies used as estimates of the phrase probabilities.

The target language model is typically the standard n -gram language model:

$$h_{\text{LM}}(f_1^J, e_1^I) = \log \prod_{i=1}^I p(e_i | e_{i-n+1}^{i-1}) \quad (4)$$

Finally, two length penalties (counting words and phrases, respectively) are included as additional features.

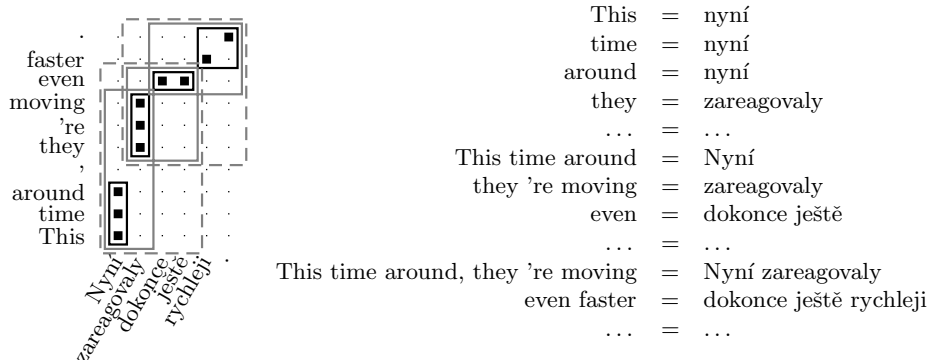


Fig. 1. Sample word alignment and sample phrases consistent with it (not all consistent phrases have been marked)

1.2 Data Description

The Prague Czech-English Dependency Corpus v. 1.0 (PCEDT [5]) consists of half of the Wall Street Journal part of Penn Treebank [6] translated sentence by sentence to Czech. Basic statistics about the training part of the PCEDT are given in Table 1. The PCEDT contains also separate development and evaluation parts (Devtest and Etest), each containing about 250 sentences with 4 independent re-translations back to English. Due to the original English source and nature of translation (sentence by sentence), the Czech sentences might be actually restricted in grammar and might not exhibit all complex word order phenomena as an independent Czech text would do. For a completely fair comparison, when the PCEDT is used to evaluate MT from English to Czech, we would need the reference translations for this direction, too.

Table 1 documents the morphological richness of Czech: the vocabulary size of Czech word forms is nearly twice as large as the vocabulary of English. If the text

Table 1. Characteristics of the Prague Czech-English Dependency Treebank 1.0

		Czech	English
	Sentences	21,141	
	Running Words	475,719	494,349
	Running Words without Punct.	404,523	439,304
Baseline (word forms)	Vocabulary	57,085	30,770
<i>Produkce malých vozů se více než ztrojnásobila .</i>	Singletons	31,458	14,637
Lemmas	Vocabulary	28,007	25,000
<i>produkce malý vůz se hodně než-2 ztrojnásobit .</i>	Singletons	13,009	11,873
Lemmas + Singletons backed off with POS	Vocabulary	15,041	13,150
<i>produkce malý vůz se hodně než-2 UNK-verb .</i>	Singletons	12	2
Stemming	Vocabulary	17,393	13,525
<i>Prod malý vozů se více než ztro .</i>	Singletons	6,347	4,846

is automatically lemmatized (this type of annotation is ready in the PCEDT), the disproportion almost disappears. In order to reduce the vocabulary size by another half, we replace all tokens appearing only once with their part of speech. A simple stemming technique (use first 4 characters of each word) gives us a the vocabulary size somewhere between lemmatization and lemmatization with singletons.

2 Techniques Improving Translation Quality

We evaluate the translation quality with the standard implementation of BLEU [7], as available for NIST evaluation³ and with the default setting (4-grams, case insensitive). An independent implementation of the BLEU metric was used to estimate confidence intervals for all the scores. Statistically significant improvements over the respective baseline are marked with a star in all the following tables.

We use the designated development and evaluation sections of the PCEDT. Results on the development section are reported with the default weights for all model parameters, results on the test set are reported after some tuning of model parameters (optimization) using the development data.

2.1 Preprocessing Czech and Choosing Type of Word Alignment

We use the GIZA++ toolkit [8] to learn word alignments. The toolkit is capable of guessing 1-n alignments (many target words are assigned to one source word). Typically, it is used twice to obtain alignments in both directions and there are two common ways to join them to a symmetric alignment: either the two directions are combined using intersection or using union.⁴ See Figure 1 for a sample union alignment.

Table 2. Translation quality and alignment error rate depending on alignment symmetrization and data preprocessing

	BLEU (ETest)		Alignment Error Rate	
	Intersection	Union	Intersection	Union
Baseline (word forms)	0.282	0.298	27.4	25.5
Stemming	-	0.306	-	-
Lemmas	0.298	0.320*	15.0	17.2
Lemmas + singletons	0.308*	0.319*	14.6	17.4

In addition to the choice of a symmetrization method, we can also employ various techniques of preprocessing tokens in the training corpus. The basic options are illustrated in Table 1: either the tokens are kept as word forms, lemmatized or

³ <http://www.nist.gov/speech/tests/mt/resources/scoring.htm>, we used the version 11b.

⁴ For other symmetrization techniques see [9].

simply stemmed. It should be noted that the preprocessing is used for estimating word alignments only. Phrases consistent with the alignment are extracted using original word forms. The translation process thus remains unchanged, i.e. we translate from source word forms to target word forms directly, only the phrase table is estimated more reliably thanks to the better alignment.

Table 2 summarizes the improvements of translation quality depending on the type of symmetrization used (intersection or union) and on the preprocessing of parallel text for alignment. We report also the alignment error rates (AER) evaluated against manually annotated alignments. See [10] for more details on the AER measurements and manual annotation. The data are directly comparable because we share the set of sentences used for the evaluation.

Similarly to [10], we observe that the reduction of vocabulary size by lemmatization significantly improves not only AER but also translation quality. (Nearly the same level of BLEU is achieved using simple stemming.). The type of symmetrization on the other hand comes out differently: based on the AER, one would choose intersection, but it leads to significantly worse translation compared to the union.

2.2 Handling Numbers

Given the type of texts in the PCEDT (economical texts), special treatment of numbers seems to pay off, see Table 3. The baseline is to treat numbers as normal tokens. To reduce the data sparseness and allow the PBT to extract phrases that correctly reorder numbers and surrounding words (mostly the dollar sign, in our case), we replace all numbers with a special symbol `_NUM`. Surprisingly, this leads to a lower performance in terms of BLEU. The best behaviour is achieved by a post-processing step to correct the typographic convention about the decimal point. As displayed in Table 3, this correction brings us some improvement, most notable on the test set (2.7% relative).

Table 3. Example of special treatment of numbers and the improvement of BLEU

	Sample input	Input to PBT	Output
Baseline	na 57,375 dolarech	na 57,375 dolarech	at 57,375 \$
Numbers	na 57,375 dolarech	na _NUM dolarech	at \$ 57,375
Numbers + Correction	na 57,375 dolarech	na _NUM dolarech	at \$ 57.375

	Devtest	Etest
Baseline	0.346	0.320
Numbers	0.341	0.309
Numbers + Correction	0.347	0.329*

2.3 Dependency-Based Corpus Expansion

Dependency syntax analysis is closely related to the notion of “sentence reduction” [11]. In short, words corresponding to leaves in the dependency structure

of the sentence can be (up to a few exceptions) removed without disrupting the grammatical correctness of the sentence. Phrase-based systems in general can learn phrase translation equivalents consisting of adjacent words only. There is a hope that a combination of these two approaches can improve translation quality, and indeed, some recent models are based on this assumption (see [12]).

We use the automatically generated dependency structure available for both Czech and English in the PCEDT to artificially expand the available training data by removing some words in the sentences. The training data for the PBT then consist of the original sentences plus a set of new sentences created by various reductions. Our method cannot be utilized off-line (before the source text to be translated is available) because there are too many possible reductions.

Given the source text, we collect all bigrams to be translated. We then scan the training data for *non-contiguous* occurrences of these bigrams (contiguous occurrences are already covered by the plain phrase extraction algorithm). For each non-contiguous occurrence we mark the two source words and then recursively add all translation equivalents (linked via word alignment) and all neighbours in both the source and the target dependency structures to satisfy some core grammatical requirements. This mainly means that at least the dependency path between all the words has to be added and some words (such as prepositions) require to add their daughters. All marked words are then printed out as a new pair of training sentences, provided that the two seed words have remained next to each other and no word has been inserted between them. (There is no point in producing a sentence pair if the words of the original bigram to be translated are not adjacent in it.)

Figure 2 illustrates the whole process of creating a new parallel phrase for the seed bigram *prověrka neukázala*. The aligned English words *check*, *n't* and *indicate* are marked first, then *seem* is added to make the English subgraph of marked words connected and finally *a*, *did* and *to* are added for grammatical reasons. In total, the new phrase *prověrka neukázala = a check did n't seem to indicate* is produced.

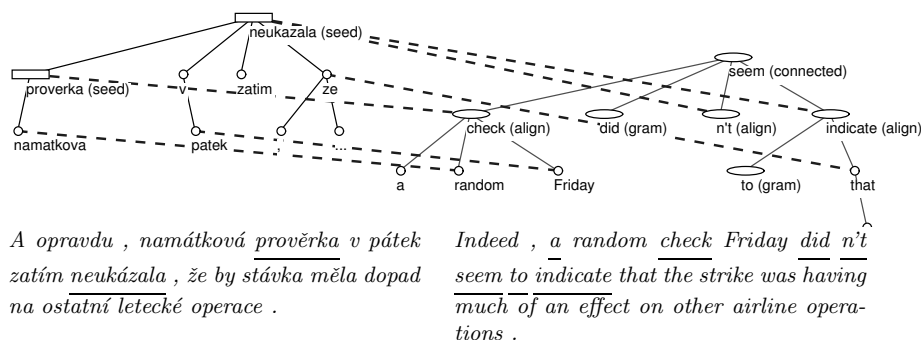


Fig. 2. Excerpts from dependency trees of word-aligned sentences illustrating dependency-based corpus expansion

Table 4 summarizes the BLEU scores on the development and evaluation set for various training corpus sizes. We have to conclude that the contribution of dependency-based corpus expansion is not statistically significant. We believe that the main reason for the failure might be inherently implied by the distributional properties of language expressions: if two words tend to depend on each other, they also tend to occur adjacently (and are thus captured by plain phrases). In other words, the situation where our algorithm can apply is rather exceptional. Indeed, only about a thousand distinct translation pairs were generated from the 20k corpus. Moreover, random errors from various sources (errors in the training sentences as such, errors in automatic parsing or limitations of the core grammatical requirements applied in our algorithm) lead to wrong translation pairs that are then inevitably suppressed by the language model.

2.4 Additional Data Sources

As documented in Table 4, doubling the parallel corpus size increases BLEU by about 0.02 to 0.04. A similar observation was reported also by [13] for Arabic-to-English.

Table 5 reports scores achieved using additional training data. Adding out-of-domain parallel texts (a collection of electronically available books) proves to bring another improvement of about 0.02 (less significant on the evaluation set). For alignment training with this additional parallel data, we did not use full lemmatization but only a simple stemming mechanism (keeping first 4 characters of words).

In a separate experiment, we employed a bigger target language model based on a monolingual corpus of the Wall Street Journal (see [3]) instead of a LM derived from the parallel texts only. As we see, adding an in-domain LM can actually serve better than adding parallel texts.

The best results we are able to achieve combine the two additional data sources: for the extraction of translation phrases, we use all parallel texts available, but only the in-domain LM is used.

Table 4. Dependency-based corpus expansion does not improve translation quality

Training sentences	Devtest			Etest		
	5k	10k	20k	5k	10k	20k
Baseline	0.275	0.316	0.346	0.254	0.284	0.320
Expanded Corpus	0.274	0.319	0.345	0.250	0.280	0.323

Table 5. Impact of additional data sources

	Devtest	Etest
Baseline: 20k sentences	0.346	0.320
20k + 85k out-of-domain sentences	0.366*	0.324
20k sentences, bigger in-domain LM	0.379*	0.337*
20k + 85k out-of-domain sentences, bigger in-domain LM	0.409*	0.370*

2.5 Finding and Fixing Clear Problems

Figure 3 illustrates our method for finding most apparent translation “errors”. We compare the set of bigrams of the hypothesis and the four reference translations on the development data. The BLEU metric penalizes our hypothesis if it contains an n-gram not present in any of the hypothesis (*superfluous n-gram*). On the contrary, the hypothesis is suspicious if it does not contain n-grams that all or most reference translations do (*missing n-gram*).

We see that the training data and the reference translations follow different typographic conventions, for instance the system tends to produce “’ ’ .” but the reference translations expect “. ”. Unfortunately, BLEU is sensitive to these differences (see also [14] for suggestions on improving correlation between BLEU and human judgements). Table 6 documents that four simple string-replacement rules inspired by the top missing and superfluous bigrams improve BLEU scores by 1.5% to 5% relative both for small and full training corpus size. The biggest improvement is observed on the development set and the positive effect is slightly reduced on the evaluation set if model parameters are optimized properly.

Top missing bigrams:		Top superfluous bigrams:	
19 , "	12 " said	26 , ' '	18 ' ' .
12 of the	10 Free Europe	14 " said	12 , which
10 Radio Free	7 . "	11 Svobodn Evropa	8 , when
6 L.J. Hooker	6 United States	8 the state	7 , who
6 in the	6 the United	7 J. Hooker	7 L. J.
6 the strike	5 " We	7 company GM	7 firm Hooker

Fig. 3. Summary of most frequent causes of loss in BLEU score

3 Summary and Related Work

Table 7 compares our best results with the results given in [3] for DBMT (Dependency-Based Machine Translation system by [3]) and ReWrite (word-based statistical MT by [15]). To the best of our knowledge, there are no other reports on the evaluation of Czech-to-English MT quality. The scores are directly comparable, because we use the same training data, language model and development and evaluation sets. Throughout this paper, BLEU scores are based

Table 6. Four patterns fixing typographic conventions significantly improve BLEU

	Devtest		Etest	
	Baseline	Fixed	Baseline	Fixed
5k sentences	0.275	0.291*	0.254	0.256
20k sentences	0.346	0.363*	0.320	0.325
20k sentences + bigger LM	0.379	0.397*	0.337	0.342

Source

Konsorcium soukromých investorů fungující jako LJH Funding Co. sdělilo, že dalo nabídku za 409 milionů dolarů v hotovosti na většinu holdingů v oblasti realit a nákupních center firmy L. J. Hooker Corp. Tato 409 milionová nabídka zahrnuje také odhadovaných 300 milionů dolarů v zaručených závazcích na tyto nemovitosti, jak uvádí nabízející strana. Skupinu vede Jay Shidler, výkonný ředitel Shidler Investment Corp. na Honolulu, a A. Boyd Simpson, výkonný ředitel Simpson Organization Inc. v Atlantě. Firma pana Shidlera se specializuje na investice do obchodních realit a chlubí se majetkem v hodnotě 1 miliardy dolarů; pan Simpson je developer a bývalý vedoucí pracovník ve firmě L. J. Hooker. "Aktiva jsou dobrá, ale vyžadují více peněz a řízení" než může L. J. Hooker v současné situaci nabídnout, řekl pan Simpson v jednom rozhovoru. "Filozofie firmy Hooker byla postavit a prodat. My chceme postavit a ponechat si. L. J. Hooker se sídlem v Atlantě funguje s ochranou proti svým věřitelům podle kapitoly 11 amerického zákona o bankrotu.

Output of the system

The private investors working as LJH Funding Co. said it could offer for \$409 million in cash for most holding in the area real-estate and shopping-center firm L.J. Hooker Corp. The 409 million offer includes also an estimated \$300 million of secured obligations on those real estate, according union-bidder party. Leading Jay Shidler, executive director Shidler Investment Corp. to Honolulu, and A. Boyd Simpson, executive director of Simpson Organization Inc. in Atlanta. The firm Mr. Shidlera specializes in investment in commercial real-estate and boasts property \$1 billion ; Mr. Simpson is the developer and former executive at the company L.J. Hooker. " Assets are good, but require more money and manage " than can L.J. Hooker in the current situation offer, said Mr. Simpson in an interview ". Philosophy Hooker's was to build and sell. We want to build and maintain. L.J. Hooker, based in Atlanta works with protection against their creditors under Chapter 11 of the United States bankruptcy law.

One of the four reference translations

A group of private investors operating under the name LJH Funding Co. has announced that they have submitted a bid of \$409 million in cash for the majority of L.J. Hooker Corp. holdings in the field of real-estate and shopping centers. This offer of \$409 million also includes a estimated \$300 million in secured bonds of this real estate, claimed the bidder. The leaders of the group are Jay Shidler, executive director of Shidler Investment Corp. in Honolulu, and A.Boyd Simpson, executive director of Simpson Organization Inc. in Atlanta. Shidler's company specializes in investments in commercial real estate, and boasts assets of \$1 billion; Simpson is a developer and former chief executive of L.J. Hooker. "The assets are sound but they require more money and management" than L.J. Hooker can offer at present, said Simpson in an interview. Hooker's philosophy has been to build and sell. We want to build and keep. L.J. Hooker, based in Atlanta, is protected against its creditors pursuant to chapter 11 of the American bankruptcy act.

Fig. 4. Sample translations using more parallel texts and the bigger in-domain language model

on four re-translations of the Czech text, in [3], the original English text is used as the fifth reference and the average over 4-reference scores (always leaving one reference out) is reported. For the purposes of comparison in Table 7, we evaluated our methods using the same averaging technique, too.

Table 7. Best results of PBT compared to other approaches

	Average over 5 refs.		4 refs. only	
	Devtest	Etest	Devtest	Etest
DBMT with parser I, no LM	0.1857	0.1634	-	-
DBMT with parser II, no LM	0.1916	0.1705	-	-
GIZA++ & ReWrite, bigger LM	0.2222	0.2017	-	-
PBT, no additional LM	0.387±0.015	0.348±0.013	0.363	0.325
PBT, bigger LM	0.413±0.012	0.364±0.013	0.397	0.342
PBT, more parallel texts, bigger LM	0.423±0.011	0.381±0.008	0.410	0.368

The results reported for PBT are based on union alignments of lemmatized training texts and the final hypotheses are typographically corrected as described in section 2.5. The language model used for our experiments is trained either on the English side of parallel texts only (“no additional LM”) or on a large monolingual corpus of Wall Street Journal, same as used in [3] (“bigger LM”). The translation of a few sentences of the Devtest are given in Figure 4.

4 Conclusion

We described several experiments with Czech-to-English phrase-based machine translation. Employing a technique of handling morphological richness of Czech is crucial, be it simple stemming or full lemmatization. The type of alignment used for phrase extraction has to be chosen carefully, too. Moreover, the alignment has to be selected on the basis of an end-to-end translation quality metric, because comparing alignments against human-annotated data leads to a sub-optimal selection.

We also experimented with rule-based handling of numbers and with a novel technique for artificial expansion of training corpus using dependency structures of the sentences.

We confirm that adding more training data improves translation quality, but it is documented that the best results are achieved if we use out-of-domain data to extract phrases only and keep the target language model in-domain. We also suggest a simple technique to find the most apparent causes of a loss in the BLEU score.

In conclusion, phrase-based statistical MT from Czech to English performs well, despite the expectations arising from linguistic knowledge about the properties of Czech. The system we experimented with is currently the best performing MT evaluated on this language pair.

Acknowledgement

The work on this experiment was partially supported by the grants GAAV ČR 1ET201120505, GAUK 351/2005 and Collegium Informaticum GAČR 201/05/H014.

References

1. Hajič, J.: Complex Corpus Annotation: The Prague Dependency Treebank. In Šimková, M., ed.: *Insight into Slovak and Czech Corpus Linguistics*, Bratislava, Slovakia, Veda, vydavateľstvo SAV (2005) 54–73
2. Sgall, P., Hajičová, E., Panevová, J.: *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Academia/Reidel Publishing Company, Prague, Czech Republic/Dordrecht, Netherlands (1986)
3. Čmejrek, M., Cuřín, J., Havelka, J.: Czech-English Dependency-based Machine Translation. In: *EACL 2003 Proceedings of the Conference, Association for Computational Linguistics (2003)* 83–90
4. Zens, R., Bender, O., Hasan, S., Khadivi, S., Matusov, E., Xu, J., Zhang, Y., Ney, H.: The RWTH Phrase-based Statistical Machine Translation System. In: *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, Pittsburgh, PA (2005) 155–162
5. Čmejrek, M., Cuřín, J., Havelka, J., Hajič, J., Kuboň, V.: Prague Czech-English Dependency Treebank: Syntactically Annotated Resources for Machine Translation. In: *Proceedings of LREC 2004, Lisbon (2004)*
6. Linguistic Data Consortium: Penn Treebank 3, LDC99T42 (1999)
7. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a Method for Automatic Evaluation of Machine Translation. In: *ACL 2002, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania (2002) 311–318
8. Och, F.J., Ney, H.: A systematic comparison of various statistical alignment models. *Computational Linguistics* **29**(1) (2003) 19–51
9. Matusov, E., Zens, R., Ney, H.: Symmetric Word Alignments for Statistical Machine Translation. In: *Proceedings of COLING 2004, Geneva, Switzerland (2004)* 219–225
10. Bojar, O., Prokopová, M.: Czech-English Word Alignment. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, ELRA (2006) (in print).
11. Lopatková, M., Plátek, M., Kuboň, V.: Modeling syntax of Free Word-Order Languages: Dependency Analysis By Reduction. In Matoušek, V., Mautner, P., Pavelka, T., eds.: *Text, Speech and Dialogue: 8th International Conference, TSD 2005, Karlovy Vary, Czech Republic, September 12-15, 2005. Proceedings. Volume LNAI 3658.*, Springer Verlag (2005) 140–147
12. Chiang, D.: A hierarchical phrase-based model for statistical machine translation. In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, Ann Arbor, Michigan, Association for Computational Linguistics (2005) 263–270
13. Och, F.J.: *Statistical Machine Translation: Foundations and Recent Advances*. Tutorial at MT Summit 2005 (2005)
14. Leusch, G., Ueffing, N., Vilar, D., Ney, H.: Preprocessing and Normalization for Automatic Evaluation of Machine Translation. In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, Ann Arbor, Michigan, Association for Computational Linguistics (2005) 17–24
15. Germann, U.: Greedy decoding for statistical machine translation in almost linear time. In: *HLT-NAACL (2003)*