Improving Statistical MT through Morphological Analysis

Sharon Goldwater
Dept. of Cognitive and Linguistic Sciences
Brown University
sharon_goldwater@brown.edu

David McClosky
Dept. of Computer Science
Brown University
dmcc@cs.brown.edu

Abstract

In statistical machine translation, estimating word-to-word alignment probabilities for the translation model can be difficult due to the problem of sparse data: most words in a given corpus occur at most a handful of times. With a highly inflected language such as Czech, this problem can be particularly severe. In addition, much of the morphological variation seen in Czech words is not reflected in either the morphology or syntax of a language like English. In this work, we show that using morphological analysis to modify the Czech input can improve a Czech-English machine translation system. We investigate several different methods of incorporating morphological information, and show that a system that combines these methods yields the best results. Our final system achieves a BLEU score of .333, as compared to .270 for the baseline word-to-word system.

1 Introduction

In a statistical machine translation task, the goal is to find the most probable translation of some foreign language text f into the desired language e. That is, the system seeks to maximize $P(e|f)$. Rather than maximizing $P(e|f)$ directly, the standard noisy channel approach to translation uses Bayes inversion to split the problem into two separate parts:

$$\underset{e}{\operatorname{argmax}} \mathrm{P}(e|f) = \underset{e}{\operatorname{argmax}} \mathrm{P}(e)\mathrm{P}(f|e) \qquad (1)$$

where $P(e)$ is known as the *language model* and $P(f|e)$ is known as the *translation model*. The limiting factor in machine translation is usually the quality of the translation model, since the monolingual resources needed for training the language model are generally more available than the parallel corpora needed for training the translation model.

Due to the difficulty in obtaining large parallel corpora, sparse data is a serious issue when estimating the parameters of the translation model. This problem is compounded when one or both of the languages involved is a highly inflected language. In this paper, we present a series of experiments suggesting that morphological analysis can be used to reduce data sparseness and increase similarity between languages, thus improving the quality of machine translation for highly inflected languages. Our work is on a language pair in which the input language (Czech) is highly inflected, and the output language (English) is not. We discuss in Section 5 how our methods might be generalized to pairs where both languages are highly inflected.

The plan of this paper is as follows: In Section 2, we review previous work on using morphological analysis for statistical machine translation. In Section 3, we describe several methods for utilizing morphological information in a statistical translation model. Section 4 presents the results of our experiments using these methods. Sections 5 and 6 discuss the results of our experiments and conclude the paper.

2 Previous Work

Until recently, most machine translation projects involved translating between languages with relatively little morphological structure. Nevertheless, a few research projects have investigated the use of morphology to improve translation quality. Niessen and Ney (2000; 2004) report work on German-English translation, where they investigate various types of morphosyntactic restructuring, including merging German verbs with their detached prefixes, annotating a handful of frequent ambiguous German words with POS tags, combining idiomatic multi-word expressions into single words, and undoing question inversion and *do*-insertion in both German and English. In addition, Niessen and Ney (2004) decompose German words into a hierarchical representation using lemmas and morphological tags, and use a MaxEnt model to combine the different levels of representation in the translation model. The results from these papers indicate that on corpus sizes up to 60,000 parallel sentences, the restructuring operations yielded a large improvement in translation quality, but the morphological decomposition provided only a slight additional benefit. However, since German is not as morphologically complex as Czech, we might expect a larger benefit from morphological analysis in Czech.

Another project utilizing morphological analysis for statistical machine translation is described by Lee (2004). Lee's system for Arabic-English translation takes as input POS-tagged English and Arabic text, where the Arabic words have been pre-segmented into stems and affixes. The system performs an initial alignment of the Arabic morphemes to the English words. Based on the consistency of the English POS tag that each Arabic morpheme aligns to, the system determines whether to keep that morpheme as a separate item, merge it back onto the stem, or delete it altogether. In addition, multiple occurrences of the determiner *Al* within a single Arabic noun phrase are deleted (i.e. only one occurrence is allowed). Using a phrase-based translation model, Lee found that *Al*-deletion was more helpful than the rest of the morphological analysis. Also, *Al*-deletion helped for training corpora up to 3.3 million sentences, but the other morphological analysis helped only on the smaller corpus sizes (up to 350,000 parallel sentences). This result is consistent with anecdotal evidence suggesting that morphological analysis becomes less helpful as corpus sizes increase. However, since parallel corpora of hundreds of thousands of sentences or more are often difficult to obtain, it would still be worthwhile to develop a method for improving systems trained on smaller corpora.

Previous results on Czech-English machine translation suggest that morphological analysis may be quite productive for this highly inflected language where there is only a small amount of closely translated material. Čmejrek et al. (2003), while not focusing on the use of morphology, give results indicating that lemmatization of the Czech input improves BLEU score relative to baseline. These results support the earlier findings of Al-Onaizan et al. (1999), who used subjective scoring measures. Al-Onaizan et al. measured translation accuracy not only for lemmatized input, but for an input form they refer to as *Czech'*. Czech' is intended to capture many of the morphological distinctions of English, while discarding those distinctions that are Czech-specific. The Czech' input was created by distinguishing the Czech lemmas for singular and plural nouns, different verb tenses, and various inflections on pronouns. Artificial words were also added automatically in cases where syntactic information in the Czech parse trees indicated that articles, pronouns, or prepositions might be expected in English. The transformation to Czech' provided a small additional increase in translation quality over basic lemmatization.

The experiments described here are similar to those performed by Al-Onaizan et al. (1999), but there are several important differences. First, we use no syntactic analysis of the Czech input. Our intent is to determine how much can be gained by a purely morphological approach to translation. Second, we present some experiments in which we modify the translation model itself to take advantage of morphological information, rather than simply transforming the input. Finally, our use of BLEU scores rather than subjective measurements allows us to perform more detailed evaluation. We examine the effects of each type of morphological information separately.

3 Morphology for MT

Morphological variations in Czech are reflected in several different ways in English. In some cases, such as verb past tenses or noun plurals, morphological distinctions found in Czech are also found in English. In other instances, English may use function words to express a meaning that occurs as a morphological variant in Czech. For example, genitive case marking can often be translated as *of* and instrumental case as *by* or *with*. In still other instances, morphological distinctions made in Czech are either completely absent in English (e.g. gender on common nouns) or are reflected in English syntax (e.g. many case markings). Handling these correspondences between morphology and syntax requires analysis above the lexical level and is therefore beyond the scope of this paper. However, morphological analysis of the Czech input can potentially be used to improve the translation model by exploiting the other types of correspondences we have mentioned.

Before we describe how this can be done, it is important to clarify the kind of morphological analysis we assume in our input. Our data comes from the Prague Czech-English Dependency Treebank (PCEDT) (Hajič, 1998; Čmejrek et al., 2004), the Czech portion of which has been fully annotated with morphological information. Each Czech word in the corpus is associated with an analysis containing the word's lemma and a sequence of morphological

```
Pro/pro/RR--4----------
někoho/někdo/PZM-4----------
by/být/Vc-X---3-------
její/jeho/PSZS1FS3-------
provedení/provedení/NNNS4-----A----
mělo/mít/VpNS---XR-AA---
smysl/smysl/NNIS4-----A----
././Z:-------------
```

Figure 1: A sentence from the PCEDT corpus. Each token is followed by its lemma and a string giving the values of up to 15 morphological tags. Dashes indicates tags that are not applicable for a particular token. This sentence corresponds to the English sentence *It would make sense for somebody to do it.*

tags. These tags provide values along several morphological dimensions, such as part of speech, gender, number, tense, and negation. There are a total of 15 dimensions along which words may be characterized, although most words have a number of dimensions unspecified. An example sentence from the Czech corpus is shown in Figure 1.

In what follows, we describe four different ways that the Czech lemma and tag information can be used to modify the parameters of the translation model. The first three of these are similar to the work of Al-Onaizan et al. (1999) and involve transformations to the input data only. The assumptions underlying the word alignment model $P(f_j|e_i)$ (where f_j and e_i are individual words in an aligned sentence pair) are maintained. The fourth method of incorporating morphological information is novel and changes the alignment model itself.

3.1 Lemmas

A very simple way to modify the input data using morphological information is by replacing each wordform with its associated lemma (see Figure 2). Based on previous results (Al-Onaizan et al., 1999; Čmejrek et al., 2003), we expected that this transformation would lead to an improvement in translation quality due to reduction of data sparseness. However, since lemmatization does remove some useful information from the Czech wordforms, we also tried two alternative lemmatization schemes. First, we tried lemmatizing only certain parts of speech, leaving other parts of speech alone. We reasoned that nouns, verbs, and pronouns all carry inflectional morphology in English, so by lemmatizing only the other parts of speech, we might retain some of the benefits of full lemmatization without losing as much information. We also tried lemmatizing all parts of speech except pronouns, which are very common and therefore should be less affected by sparse data problems.

As a second alternative to full lemmatization, we experimented with lemmatizing only the less frequent wordforms in the corpus. This allows the translation system to use the full wordform information from more frequent forms, where sparse data is less of a problem.

To determine whether knowledge of lemmas was actually necessary, we compared lemmatization with word truncation. We truncated each wordform in the data after a fixed number of characters, as suggested by Och (2005).

3.2 Pseudowords

As discussed earlier, much of the information encoded in Czech morphology is encoded as function words in English. One way to reintroduce some of the information lost during Czech lemmatization is by using some of the morphological tags to add extra "words" to the Czech input. In many cases, these pseudowords will also increase the correspondence of English function words to items in the Czech input. In our system, each pseudoword encodes a single morphological tag (feature/value pair), such as PER_1 ('first person') or TEN_F ('future tense'). Figure 2 shows a Czech input sentence after generating pseudowords for the person feature on verbs.

We expected that the class of tags most likely to be useful as pseudowords would be the person tags, because Czech is a pro-drop language. Using the person tags as pseudowords should simulate the existence of pronouns for the English pronouns to align to. We also expected that negation (which is expressed on verbs in Czech) would be a useful pseudoword, and that case markings might also be helpful since they sometimes correspond to prepositions in English, such as *of*, *with*, or *to*.

3.3 Modified Lemmas

In some cases, such as the past tense, Czech morphology is likely to correspond not to a function word in English, but rather to English inflectional morphology. In order to capture this kind of phenomenon, we experimented with concatenating the Czech morphological tags onto their lemmas instead of inserting them as separate input tokens. See Figure 2 for an example. This concatenation creates distinctions between some lemmas, which will ideally correspond to morphological distinctions made in English. Although this transformation splits the Czech data (relative to pure lemmatization), it still suppresses many of the distinctions made in the full Czech wordforms. We expected that number mark-

Words:	`Pro někoho by její provedení mělo smysl .`
Lemmas:	`pro někdo být jeho provedení mít smysl .`
Lemmas+Pseudowords:	`pro někdo být PER_3 jeho provedení mít PER_X smysl .`
Modified Lemmas:	`pro někdo být+PER_3 jeho provedení mít+PER_X smysl .`

Figure 2: Various transformations of the Czech sentence from Figure 1. The pseudowords and modified lemmas encode the verb person feature, with the values 3 (third person) and X ("any" person).

ing on nouns and tense marking on verbs would be the tags best treated in this way.

3.4 Morphemes

Our final set of experiments used the same input format as the Modified Lemma experiments. However, in this set of experiments, we changed the model used to calculate the word-to-word alignment probabilities. In the standard system, the alignment model parameters $P(f_j|e_i)$ are found using maximum likelihood estimation based on the expected number of times f_j aligns to e_i in the parallel corpus. Our new model assumes a compositional structure for f_j, so that $f_j = f_{j0} \ldots f_{jK}$, where f_{j0} is the lemma of f_j, and $f_{j1} \ldots f_{jK}$ are morphemes generated from the tags associated with f_j. We assume that every word contains exactly K morphemes, and that the kth morpheme in each word is used to encode the value for the kth class of morphological tag, where the classes (e.g. person or tense) are assigned an ordering beforehand. f_{jk} is assigned a null value if the value of the kth tag class is unspecified for f_j.

Given this decomposition of words into morphemes, and a generative model in which each morpheme in f_j is generated independently conditioned on e_i, we have

$$\mathrm{P}(f_j|e_i) = \prod_{k=0}^{K} \mathrm{P}(f_{jk}|e_i) \tag{2}$$

We can now estimate $\mathrm{P}(f_j|e_i)$ using a slightly modified version of the standard EM algorithm for learning alignment probabilities. During the E step, we calculate the expected alignment counts between Czech morphemes and English words based on the current word alignments, and revise our estimate of $\mathrm{P}(f_j|e_i)$ using Equation 2. The M step of the algorithm remains the same.

The morpheme-based model in Equation 2 is similar to the modified lemma model in that it removes much of the differentiation between Czech wordforms, but leaves the differences that are most likely to appear as inflection on English words. However, it also performs an additional smoothing function. The model assumes that, in the absence of other information, an English word that has aligned mostly to Czech words with a particular morphological tag is more likely to align to another word with this tag than to a Czech word with a different tag. For example, an English word aligned to mostly past tense forms is more likely to align to another past tense form than to a present or future tense form.

4 Experiments

In order to evaluate the effectiveness of the techniques described in the previous section, we ran a number of experiments using data from the PCEDT corpus. The English portion of this corpus (used to train the language model) contains the same material as the Penn WSJ corpus, but with a different division into training, development, and test sets. About 250 sentences each for development and test were translated once into Czech and then back into English by five different translators. These translations are used to calculate BLEU scores. The remainder of the corpus (about 50,000 sentences) is used for training. About 21,000 of the training sentences have been translated into Czech and morphologically annotated for use as a parallel corpus.

Some statistics on the parallel corpus are shown in the graph in Figure 3. This graph illustrates the sparse data problem in Czech that our morphological analysis is intended to address. Although the number of infrequently occurring lemmas is about the same in both English and Czech, the number of infrequently occurring inflected wordforms is approximately twice as high in Czech.[1]

For all of our experiments, we used the same language model, trained with the CMU Statistical Language Modelling Toolkit (Clarkson and Rosenfeld, 1997). Our translation models were trained using GIZA++ (Och and Ney, 2003), which we modi-

[1]Although we did not use it for the experiments in this paper, the PCEDT corpus does contain lemma information for the English data. There is a slight discrepancy between the English and Czech data in the lemma information for pronouns, in that English pronouns (including accusitive, possessive, and other forms) are assigned themselves as lemmas, whereas Czech pronouns are reduced to uninflected forms. Given that pronouns generally have many tokens, this discrepancy should not affect the data in Figure 3.

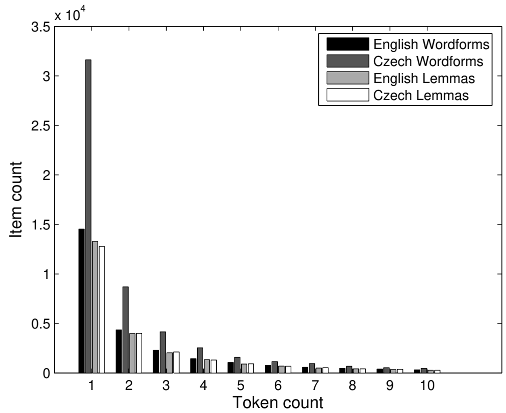

Figure 3: The number of items (full wordforms or lemmas) y appearing in the parallel corpus with a token count of x.

fied as necessary for the morpheme-based experiments. We used the ISI ReWrite Decoder (Marcu and Germann, 2005) for producing translations. Before beginning our experiments, we obtained a baseline BLEU score by training a standard word-to-word translation model. Our baseline results indicate that the test set for this corpus is considerably more difficult than the development set: word-to-word scores were .311 (development) and .270 (test).

4.1 Lemmas

As Figure 3 shows, lemmatization of the Czech corpus cuts the number of unique items by more than half, and the number of items with no more than ten occurrences by nearly half. The lemmatization BLEU scores in Table 1 indicate that this has a large impact on the quality of translation. As expected, full lemmatization performed better than word-to-word translation, with an an improvement of about .04 in the development set BLEU score and .03 in the test set. (In this and the following experiments, BLEU score differences of .009 or more are significant at the .05 level.) Experiments on the development set showed that leaving certain parts of speech unlemmatized did not improve results, but lemmatizing only low-frequency words did. A frequency cutoff of 50 worked best on the development set (i.e. only words with frequency less than 50 were lemmatized). Despite the improvement on the development set, using this cutoff with the test set yielded only a non-significant improvement over full lemmatization.

The results of these lemmatization experiments support the argument that lemmatization improves translation quality by reducing data sparseness, but also removes potentially useful information. Our results suggest that lemmatizing only infrequent words may, in some cases, work better than lemmatizing all words.

	Dev	Test
word-to-word	.311	.270
lemmatize all	.355	.299
except Pro	.350	
except Pro, V, N	.346	
lemmatize $n < 50$	.370	.306
truncate all	.353	.283

Table 1: BLEU scores for the word-to-word baseline, lemmatization, and word truncation experiments.

As Table 1 indicates, it is possible to get some of the benefits of lemmatization without using any morphological knowledge at all. For both dev and test sets, truncating words to 6 characters (the best length on the dev set) provided a significant improvement over word-to-word translation, but was also significantly worse than the best lemmatization scores. Changing the frequency cutoff for truncation did not produce any significant differences in the BLEU score.

4.2 Pseudowords

Results for the pseudoword experiments on the development set are shown in the first column of Table 2. Note that in these (and the following) experiments, we treated all words the same way regardless of their frequency, so the effects of adding morphological information are in comparison to the full lemmatization scheme. In most of our experiments, we added morphological information for only a single class of tags at a time in order to determine the effects of each class individually. The classes we used were verb person (PER), verb tense (TEN), noun number (NUM), noun case (CASE), and negation (NEG).

Most of the results of the pseudoword experiments confirm our expectations. Adding the verb person tags was helpful, and examination of the alignments revealed that they did indeed align to English pronouns with high probability. The noun number tags did not help, since plurality is expressed as an affix in English. Negation tags helped slightly, though the improvement was not significant. This is probably because negation tags are relatively infrequent, as can be seen in Table 3. The addition of pseudowords for case did not yield an improvement, probably because these pseudowords were so frequent. The additional ambiguity caused by so many extra words likely overwhelmed any positive effect.

A somewhat puzzling result is the behavior of the

Tag type	Pseudo	Mod-Lem	Morph
PER	.365	.356	.356
TEN	.365	.361	.364
PER,TEN	.355	.362	.355
NUM	.354	.367	.361
CASE	.353	.340	.337
NEG	.357	.356	.353

Table 2: BLEU scores indicating the results of incorporating the information from different classes of morphological tags in the the experiments using pseudowords (Pseudo), modified lemmas (Mod-Lem), and morphemes (Morph). Scores are from the development set. Differences of .009 are significant ($p < .05$).

Tag class	Count	Avg/sentence
PER	49700	2.35
TEN	47744	2.26
past	22544	1.07
pres	20291	0.96
fut	1707	0.08
'any'	3202	0.15
NUM	151646	7.17
CASE	151646	7.17
NEG	3326	0.16

Table 3: Number of occurrences of each class of tags in the Czech training data.

verb tense tags. With the exception of future tense, English generally does not mark tense with an auxiliary. Yet Table 3 shows that only a very small percentage of sentences have a future tense marker, so it seems unlikely that this explains the positive effects of the tense pseudowords. In fact, we tried adding only future tense pseudowords to the lemmatized Czech data, and found that the results were no better than basic lemmatization.

The other unusual behavior we see with pseudowords is that when verb person and tense tags are combined, they seem to cancel each other out, resulting in a score that is no better than lemmatization alone. Examination of the alignments did not reveal any obvious reason for this effect.

4.3 Modified Lemmas

As shown in the second column of Table 2, the number and tense tags yield an improvement under the modified lemma transformation, while the person tags do not. Again, this confirms our predictions based on the morphology of English.

Our results using the case tags under this model actually decreased performance, but this is not surprising given that differentiating Czech lemmas based on case marking creates as much as a 7-way split of the data (there are seven cases in Czech), without adding much information that would be useful in English.

4.4 Morphemes

BLEU scores for the morpheme-based model are given in the third column of Table 2. None of the differences in scores between this model and the modified lemma model are significant, although the trend for most of the tag classes is for this model to perform slightly worse. This suggests that the type of smoothing induced by the morpheme-based model may not be as helpful as simply attempting to create Czech words that reflect the same morphological distinctions as the English words. In Section 5, we propose a generalized version of the morpheme model that might be an improvement.

4.5 Combined Model

In the experiments described so far, we used only a single method at a time of incorporating morphological information into the translation process. However, it is straightforward to combine the pseudoword method with either the modified-lemma or morpheme-based methods by using pseudowords for certain tags and attaching others to the Czech lemmas. The experiments described above allowed us to confirm our intuitions about how each class of tags should be treated under such a combined model. We then created a model using the pseudoword treatment of the person and negation tags, and the modified lemma treatment of number and tense. We did not use the case tags in this model, since they did not seem to yield an improvement in any of the three basic morphological models.

Our combined model achieved a BLEU score of .390 (development) and .333 (test), outperforming the models in all of our previous experiments.

5 Discussion

The results of our experiments provide additional support for the findings of previous researchers that using morphological analysis can improve the quality of statistical machine translation for highly-inflected languages. While human judgment is probably the best metric for evaluating translation quality, our use of the automatically-derived BLEU score allowed us to easily compare many different translation models and evaluate the effects of each one individually. We found that simple lemmatization, by significantly reducing the sparse data problem, was quite effective despite the loss of information involved. Lemmatizing the less frequent words in the corpus seemed to increase performance slightly, but these results were inconclusive. Word truncation, which requires no morphological information at all, was effective at increasing scores over the word-to-word baseline, but did not perform quite as well as lemmatization. This result conflicts with Och's (Och, 2005), and is likely due to the much smaller size of our corpus. In any case, our results suggest that lemmatization or word truncation could yield a significant improvement in the quality of translation from a highly-inflected to a less-inflected language, even when limited morphological information is available.

Our primary results concern the use of full morphological information. We found that certain tags were more useful when we treated them as discrete input words, while others provided a greater benefit when attached directly to their lemmas. The best choice of which method to use for each class of tags seems to correspond closely with how that class of information is expressed in English (either using function words or inflection). In a sense, the goal of the morphological analysis is to make the Czech input data more English-like by suppressing unnecessary morphological distinctions and expressing necessary distinctions in ways that are similar to English. This sort of procedure could be taken further by incorporating syntactic information as well, but as we stated earlier, our goal was to determine exactly how much benefit we could derive from a strictly morphological approach.

In the work we have presented, the output language (English) is low in inflection. We therefore considered it less important to perform morphological analysis on the English data. However, we expect that the work described here could be generalized to highly inflected output languages by doing morphological analysis on both the input and output languages. The most promising way to do this seems to be by extending the morpheme-based translation model in Equation 2 to incorporate morphemes in both languages, so that

$$\mathrm{P}(f_j|e_i) = \prod_{k=0}^{K} \mathrm{P}(f_{jk}|e_{ik}) \qquad (3)$$

where f_{jk} are the morphemes in the input language, and e_{ik} are the corresponding morphemes in the output language. This extended model may also prove a benefit to Czech-English translation; we are currently investigating this possibility.

In this work, we used a word-based translation system due to the availability of source code that could be modified for our morph experiments. An obvious extension to the current work would be to move to a phrase-based translation system. One advantage of phrase-based models is their ability to align phrases in one language to morphologically complex words in the other language. However, this feature still suffers from the same sparse data problems as a word-based system: if a morphologically complex word only appears a handful of times in the training corpus, the system will have difficulty determining its (phrasal or word) alignment. We expect that morphological analysis would still be helpful in this situation, at the very least because it can be used to remove distinctions that appear in only one language.

6 Conclusion

In this paper we used morphological analysis of Czech to improve a Czech-English statistical machine translation system. We have argued that this improvement was primarily due to a reduction of the sparse data problem caused by the highly inflected nature of Czech. An alternative method for reducing sparse data is to use a larger parallel corpus; however, it is often easier to obtain additional monolingual resources, such as a morphological analyzer or tagged corpus, than additional parallel data for a specific language pair. For that reason, we believe that the approach taken here is a promising one.

We have described several different ways of using morphological information for machine translation, and have shown how these can be combined to yield an improved translation model. In general, we would not expect the exact combination of techniques that yielded our best results for Czech-English to be optimal for other language pairs. Rather, we have suggested that these techniques should be combined in a way that makes the input language more similar to the output language. Although this combination will need to be determined for each language pair, the general approach outlined here should provide benefits for any MT system involving a highly inflected language.

Acknowledgements

We would like to thank Eugene Charniak and the members of BLLIP for their encouragement and helpful suggestions. This research was partially supported by NSF awards IGERT 9870676 and ITR 0085940.

References

Y. Al-Onaizan, J. Cuřin, M. Jahr, K. Knight, J. Lafferty, D. Melamed, F. Och, D. Purdy, N. Smith, and D. Yarowsky. 1999. Statistical machine translation. Final Report, JHU Summer Workshop 1999.

P. Clarkson and R. Rosenfeld. 1997. Statistical language modeling using the CMU-Cambridge Toolkit. In *Proceedings of ESCA Eurospeech*. Current version available at http://mi.eng.cam.ac.uk/~prc14/toolkit.html.

J. Hajič. 1998. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In Eva Hajičová, editor, *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, pages 12–19. Prague Karolinum, Charles University Press.

Y. Lee. 2004. Morphological analysis for statistical machine translation. In *Proceedings NAACL*.

D. Marcu and U. Germann. 2005. The ISI ReWrite Decoder 1.0.0a. Available at http://www.isi.edu/licensed-sw/rewrite-decoder/.

S. Niessen and H. Ney. 2000. Improving SMT quality with morpho-syntactic analysis. In *Proceedings of COLING*.

S. Niessen and H. Ney. 2004. Statistical machine translation with scarce resources using morpho-syntactic analysis. *Computational Linguistics*, 30(2):181–204.

F. Och and H. Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

F. Och. 2005. Statistical machine translation: The fabulous present and future. Invited talk at the Workshop on Building and Using Parallel Texts at ACL'05.

M. Čmejrek, J. Cuřín, and J. Havelka. 2003. Czech-english dependency-based machine translation. In *Proceedings of EACL*.

M. Čmejrek, J. Cuřín, J. Havelka, J. Hajič, and V. Kuboň. 2004. Prague czech-english dependecy treebank: Syntactically annotated resources for machine translation. In *4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.