

Morpho-Syntactic Analysis for Reordering in Statistical Machine Translation

Sonja Nießen and Hermann Ney

Lehrstuhl für Informatik VI
Computer Science Department
RWTH Aachen – University of Technology
D-52056 Aachen, Germany
Email: niessen@cs.rwth-aachen.de

Abstract

In the framework of statistical machine translation (SMT), correspondences between the words in the source and the target language are learned from bilingual corpora on the basis of so-called alignment models. Among other things these are meant to capture the differences in word order in different languages. In this paper we show that SMT can take advantage of the explicit introduction of some linguistic knowledge about the sentence structure in the languages under consideration. In contrast to previous publications dealing with the incorporation of morphological and syntactic information into SMT, we focus on two aspects of reordering for the language pair German and English, namely question inversion and detachable German verb prefixes. The results of systematic experiments are reported and demonstrate the applicability of the approach to both translation directions on a German-English corpus.

Keywords

Statistical Machine Translation, Reordering, Morpho-Syntactic Information, Question Inversion, Seperable Prefixes

1 Introduction

In this paper, we address the question of how morphological and syntactic analysis can help statistical machine translation (SMT). Although there has been a number of publications dealing with morphological and syntactic analysis in general and its application to machine translation in particular, there have only been a few which incorporate information from this analysis in the process of *statistical* machine translation (Brown et al., 1992; Nießen and Ney, 2000).

In our approach, we introduce transformations to both the source and target string. In our experiments the considered languages are German and English. Systematic evaluations demonstrate that linguistic knowledge can improve translation results. We concentrate on transformations which aim at “harmonizing” the word order in corresponding sentences. Our experiences with various tasks and language pairs show that difference in word order is one of the main sources of errors in machine translation, if not the most dominant problem in this field. The presentation focusses on the following aspects:

Separated German verb prefixes: Some German verbs consist of a main part and a detachable prefix which can be shifted to the end of the clause. In (Nießen and Ney, 2000), we have suggested to prepend separated prefixes to the

main part and reported results for the translation direction from German to English. Here, we extend this method and apply it to the inverse direction from English to German which implies some additional complications.

Question Inversion: In many languages, the sentence structure of interrogative sentences differs from the structure in declarative sentences. We propose a method of harmonizing the word order in both sentence types. For the language pair English and French, this procedure was suggested by (Brown et al., 1992). In contrast to them, we investigate question inversion for German, which is more complicated in comparison to French because of the less restricted word order. Unlike (Brown et al., 1992), we will report on quantitative results regarding the effect of question inversion treatment on the translation quality.

The paper is organized as follows: After reviewing the statistical approach to machine translation, we will first describe the tools we use for morphological and syntactic analysis. We will then describe the treated phenomena and present our solutions in detail. Experimental results on a bilingual German-English task will be reported. Finally, we will give an outlook on our future work.

2 Statistical Machine Translation

The goal of the translation process in statistical machine translation can be formulated as follows: Every target language string $e_1^J = e_1 \dots e_J$ is assigned a probability $Pr(e_1^J | f_1^J)$ to be an admissible translation for the given source language string $f_1^J = f_1 \dots f_J$. According to Bayes' decision rule, we have to choose the target string that maximizes the product of the target language model $Pr(e_1^J)$ and the string translation model $Pr(f_1^J | e_1^J)$.

Many existing systems for SMT (Wang and Waibel, 1997; Nießen et al., 1998; Och et al., 1999) make use of a special way of structuring the string translation model (Brown et al., 1993): The correspondence between the words in the source and the target string is described by alignments that assign target word positions to each source word position. The probability of a certain target language word to occur in the target string is assumed to depend basically only on the source words aligned to it.

The overall architecture of the statistical translation approach is depicted in Figure 1. In this figure we already anticipate the fact that we will transform the source strings in a certain manner. If necessary we also apply the inverse of these transformations on the produced output strings.

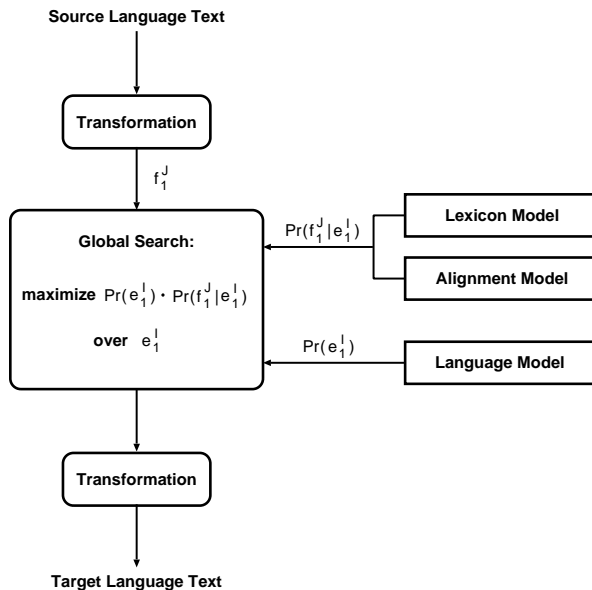


Figure 1: Architecture of the translation approach based on Bayes' decision rule.

3 Morpho-Syntactic Analysis

For lexical analysis and morphological and syntactic disambiguation, we use commercial constraint grammar parsers for German and English, respectively. For

a description of the constraint grammar approach we refer the reader to (Karlsson, 1990).

Tables 1 and 2 give examples of the information provided by these tools. These examples demonstrate that the tools can quite reliably disambiguate between different readings: For instance, they infer that the word “wollen” is a verb in the indicative present first person plural form. Without any context taken into account, “wollen” has other readings. It can even be interpreted as derived not from a verb, but from an adjective with the meaning “made of wool”.

In the cases where the tools returns more than one reading simple heuristics based on domain specific preference rules are applied. The reading “plural of Esse” for the German word form “Essen” for instance is much less likely in our domain (for a description of the domain see Section 5) than the readings “proper name of the town Essen” or the German equivalent of the English word “meal”.

lemma	tags
wir	pers pron pl 1st nom
wollen	V ind present pl 1st
nach	prep
das	art sg dat neutr
Essen	N neutr sg dat
nach	prep
Essen	N name sg dat
Esse	N fem pl dat
Essen	N neutr pl dat
Essen	N neutr sg dat
auf brechen	V seperable inf

Table 1: Sample analysis of a German sentence. Input: ‘‘Wir wollen nach dem Essen nach Essen aufbrechen’’.

lemma	tags
do	V present fin aux
we	pers pron pl 1st subject
have	V inf main
to	inf mark
arrange	V inf main
a	art sg
meet	present participle object

Table 2: Sample analysis of an English sentence. Input: ‘‘Do we have to arrange a meeting’’.

4 Transformation

As already pointed out, we use the method of transforming the input and output strings in our experiments. As a consequence, existing training and search procedures do not have to be adapted to new models incorporating the information under consideration. Transforming the training corpora implies restarting the training procedure for the parameters of the sta-

tistical models, but the algorithms themselves remain unchanged. In the following subsections we list the reordering transformations we have tested.

4.1 Separated German Verb Prefixes

Some verbs in German consist of a main part and a detachable prefix which can be shifted to the end of the clause, e.g. “losfahren” (“to leave”) in the sentence “Ich fahre morgen los.”.

For the alignment process it is often difficult to learn that one English word has to be associated with more than one word in the corresponding German sentence, namely the main part of the verb and the separated prefixes. This difficulty is more serious in the (frequent) cases, where the distance between the positions of the main and the prefix part is large.

To solve the problem of separated prefixes, we proceed as follows: We extract all separable word forms of verbs from the training corpus. The resulting list contains entries of the form `prefix|main`. For example, the entry “los|fahren” indicates, that the prefix “los” can be detached from the word form “fahren”. In all clauses containing a word matching a main part and a word matching the corresponding prefix part occurring at the end of the clause, the prefix is prepended to the beginning of the main part, as in “Ich losfahre morgen”. This is carried out for the German part of the training corpus and, in the case that German is the source language, also for the input sentences in the testing phase.

In contrast to our earlier publication (Nießen and Ney, 2000), in which we have already proposed this procedure for the translation direction from German to English, we also apply the method to the translation direction from English to German. This direction is more complicated, because we additionally need post-processing of the German output sentence to reconstruct the correct forms of the separable verbs. We use a language model rescoring approach to choose between different positionings of the verb prefix. For example, we decide whether to accept “Ich losfahre morgen”, “Ich fahre los morgen” or “Ich fahre morgen los”. For this purpose, the trigram language model scores of the original sentence and the variants with moved prefixes are computed and the best scoring translation is chosen.

4.2 Question Inversion

In German as well as in English and in many other languages, the sentence structure of interrogative sentences differs from the structure in declarative sentences in that the order of the positions of the subject and the corresponding finite verb is inverted.

We transform questions in both languages in such a way as to harmonize the word order in both sentence

types. From the perspective of statistical translation, the advantages of this method are the following:

- The standard algorithm for training the parameters of the target language model $Pr(e_1^I)$ cannot deduce the probability of a word sequence in an interrogative sentence from the corresponding declarative form. For example, from the frequency of the sequence “you would have time” in the training corpus, the language model is not able to infer the probability of the sequence “would you have time”.
- The same reasoning is valid for those statistical machine translation systems, which can learn the lexical translation probabilities of multi-word phrase pairs, like for instance the alignment template approach described in (Och et al., 1999) and used as the translation system in our experiments: Without a special treatment of question inversion, such a system would not be able to learn the translation “ist es Dir recht” for “would you mind” from the bilingual sample “(you would mind)/“es ist Dir recht””.

The procedure of harmonizing the word order of questions with the word order in declarative sentences can best be understood by looking at the examples in Figure 2: the order of the subject (including the appendant articles, adjectives etc.) and the corresponding finite verb (in English: an auxiliary verb) is inverted. In English questions supporting “do”s are removed.

For the language pair English and French, this procedure was suggested by (Brown et al., 1992), but they did not report on experimental results revealing the effect of the reordering on the translation quality. Our reordering algorithm uses the information from syntactic analysis (see Tables 1 and 2), which helps to find the subject and the corresponding finite verb in an interrogative phrase. Because of the smaller variability regarding word order in English, this information is especially explicit and reliable for English. For the cases when subject and corresponding finite verb cannot unambiguously be identified from the analysis, we implemented some heuristics which proved to be correct in most cases.

The application of the described preprocessing step on interrogative phrases in the bilingual training corpus implies the necessity of restoring the correct forms of the translations produced by the MT algorithm: In a postprocessing step the inverse reorderings are performed and if necessary, the correct forms of the supporting “do” are inserted.

5 Translation Experiments

Experiments were carried out on Verbmobil data. The Verbmobil corpus consists of spontaneously spo-

"may I take your order, sir?"	→	"I may take your order, sir?"
"do you know where the boutique is?"	→	"you know where the boutique is?"
"when would you have time for that?"	→	"when you would have time for that?"
"I mean, why should I reject it?"	→	"I mean, why I should reject it?"
"would you mind, if I come today?"	→	"you would mind, if I come today?"
"darf ich Ihre Bestellung aufnehmen?"	→	"ich darf Ihre Bestellung aufnehmen?"
"wann hätten Sie denn Zeit?"	→	"wann Sie hätten denn Zeit?"
"ich meine, warum sollte ich das tun?"	→	"ich meine, warum ich sollte das tun?"
"ist es ok, wenn ich heute komme?"	→	"es ist ok, wenn ich heute komme?"

Figure 2: Examples for the effect of reordering in questions.

ken dialogs in the appointment scheduling domain (Wahlster, 2000). The training set consists of 58 322 sentence pairs. Table 3 summarizes the characteristics of the corpus used for training the parameters of Model 4 as proposed in (Brown et al., 1993).

	English	German
no. of running words	550 213	519 790
no. of word forms	4 670	7 940
no. of singletons	1 696	3 452
singletons [%]	36	43

Table 3: Corpus statistics: Verbmobil training. Singletons are types occurring only once in training.

For testing we used the alignment template translation system described in (Och et al., 1999), which already has a reasonably good capability of performing word reordering on its own: We were interested in the question whether this systems' performance can be improved by explicit treatment of word order differences.

Tests were carried out for each of the translation directions from German to English and from English to German on about 250 sentences not contained in the training data.

5.1 Performance Metrics

We use the following evaluation criteria (Nießen et al., 2000):

SSER (subjective sentence error rate):

Each translated sentence is judged by a human examiner according to an error scale from 0.0 (semantically and syntactically correct) to 1.0 (completely wrong).

ISER (information item semantic error rate):

The test sentences are segmented into information items; for each of them, the translation candidates are assigned either "ok" or an error class. If the intended information is conveyed, the translation of an information item is considered correct,

even if there are slight syntactic errors, which do not seriously deteriorate the intelligibility.

5.2 Translation direction English to German

The test set characteristics for this translation direction are summarized in Table 4.

no. of sentences	248
no. of running words	3 040
no. of word forms	355

Table 4: Test set for English to German translation.

As Table 5 shows, both preprocessing methods as such have a positive effect on the translation quality. The further improvement by combining both methods is not very large. In a different experimental setup, where we restricted the corpus for training the model parameters to less than 10 % of the original size, the achieved improvement of the translation quality was larger. This confirms the expectation that harmonizing word order between languages and different sentence types enables a better exploitation of the bilingual training data.

Examples are given in Figure 3 for prepending of separated verb prefixes and in Figure 4 for the treatment of question inversion. In these figures as well as in the following ones, "↓" denotes the translation process itself and "↓" indicates the effect of the pre- or postprocessing.

5.3 Translation direction German to English

251 German source sentences not contained in the training corpus were translated into English. The test set characteristics are depicted in Table 6. Again, the average number of reference translations for the computation of the multi-reference word error rate is 1.6.

The results are summarized in Table 7. Prepending separated prefixes helps less when translating from German into English than the treatment of question inversion does. In comparison to the latter, the translation quality as measured by the subjective sentence

Train Size [no. of sent.]	method	SSER [%]	ISER [%]
58k (100 %)	baseline	35.2	17.3
	treat prefixes	33.7	13.3
	treat question inversion	34.0	16.3
	treat question inversion + prefixes	33.4	12.8
5k	baseline	51.3	27.2
	treat question inversion + prefixes	48.6	25.4

Table 5: Effect of preprocessing for the translation direction English to German

No treatment of separated prefixes	Prefixes prepended
“I will pick you up downtown.”	“I will pick you up downtown.”
↓	↓
“ich hole Sie in der Innenstadt.”	“ich abhole Sie in der Innenstadt.”
↓	↓
“ich hole Sie in der Innenstadt.”	“ich hole Sie in der Innenstadt ab.”

Figure 3: Example for prepending separated prefixes for the translation direction English to German.

No treatment question inversion	Question inversion treated
“where shall we stay at in Hanover?”	→ “where we shall stay at in Hanover?”
↓	↓
“wo sollen wir bleiben um in Hannover?”	“wo wir übernachten in Hannover?”
↓	↓
“wo sollen wir bleiben um in Hannover?”	“wo übernachten wir in Hannover?”

Figure 4: Example for question inversion treatment for the translation direction German to English.

no. of sentences	251
no. of running words	2 627
no. of word forms	430

Table 6: Test set for German to English translation.

error rate slightly deteriorates when combining both methods, whereas the information item error rate decreases. We conclude that the combination helps to transmit the intended information, but that this positive effect is compensated by a deteriorated syntax.

Figures 5 and 6 illustrate the effect of the pre- and postprocessings for this translation direction.

6 Conclusion and Future Work

In this paper, we have presented some methods of providing morphological and syntactic information for improving the performance of statistical machine translation. First experiments prove their applicability to realistic and complex tasks such as spontaneously spoken dialogs. The results reveal that harmonizing word order between languages and different sentence types enables a better exploitation of the bilingual training data for statistical machine translation.

We are planning to apply the approach to other tasks and other language pairs. We will also perform systematic experiments to examine the effect of the proposed preprocessing steps on the alignment quality on the training corpus and on the quality of statistical lexica.

Acknowledgement. This work was partly supported by the German Federal Ministry of Education, Science, Research and Technology under the Contract Number 01 IV 701 T4 (VERBMOBIL) and as part of the EUTRANS project by the European Community (ESPRIT project number 30268).

References

- P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, J. D. Lafferty, and R. L. Mercer. 1992. Analysis, statistical transfer, and synthesis in machine translation. In *Proc. TMI 1992: 4th Int. Conf. on Theoretical and Methodological Issues in MT*, pages 83–100, Montréal, P.Q., Canada, June.
- P. Brown, S. Della Pietra, V. Della Pietra, and R. Mercer. 1993. Mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Train Size [no. of sent.]	method	SSER [%]	ISER [%]
58k (100 %)	baseline	32.6	17.9
	treat prefixes	32.2	18.2
	treat question inversion	29.8	17.3
	treat question inversion + prefixes	30.4	16.9
5k	baseline	46.6	32.4
	treat question inversion + prefixes	45.1	30.6

Table 7: Effect of preprocessing for the translation direction German to English

No treatment of separated prefixes	Prefixes prepended
“tragen wir das ein.” ↓ “that we put a.”	→ “eintragen wir das.” ↓ “we put it down.”
“fahren wir nicht zu früh los.” ↓ “we will go not too early leave.”	→ “losfahren wir nicht zu früh.” ↓ “we will leave not too early.”

Figure 5: Examples for prepending separated prefixes for the translation direction German to English.

No treatment question inversion	Question inversion treated
“wollen Sie auch Plätze reservieren?” ↓ “do you want to also seats reserve?”	→ “Sie wollen auch Plätze reservieren?” ↓ “you also want to reserve seats?” ↓ “do you also want to reserve seats?”

Figure 6: Example for question inversion treatment for the translation direction German to English.

- F. Karlsson. 1990. Constraint grammar as a framework for parsing running text. In *Proc. COLING 1990: The 13th Int. Conf. on Computational Linguistics*, volume 3, pages 168–173, Helsinki, Finland.
- S. Nießen and H. Ney. 2000. Improving SMT quality with morpho-syntactic analysis. In *Proc. COLING 2000: The 18th Int. Conf. on Computational Linguistics*, pages 1081–1085, Saarbrücken, Germany, July.
- S. Nießen, S. Vogel, H. Ney, and C. Tillmann. 1998. A DP based search algorithm for statistical machine translation. In *Proc. COLING-ACL 1998: The 36th Annual Meeting of the Association for Computational Linguistics and the 17th Int. Conf. on Computational Linguistics*, pages 960–967, Montréal, P.Q., Canada, August.
- S. Nießen, F. J. Och, G. Leusch, and H. Ney. 2000. An evaluation tool for machine translation: Fast evaluation for MT research. In *Proc. LREC 2000: The 2nd Int. Conf. on Language Resources and Evaluation*, pages 39–45, Athens, Greece, May.
- F. J. Och, C. Tillmann, and H. Ney. 1999. Improved alignment models for statistical machine translation. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 20–28, University of Maryland, College Park, MD, June.
- W. Wahlster, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Springer Verlag: Berlin, Heidelberg, New York.
- Y.-Y. Wang and A. Waibel. 1997. Decoding algorithm in statistical translation. In *Proc. ACL-EACL 1997: The 35th Annual Meeting of the Association for Computational Linguistics - joint with EACL 1997*, pages 366–372, Madrid, Spain, July.