THE STATISTICAL APPROACH TO SPOKEN LANGUAGE TRANSLATION

Hermann Ney

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

ABSTRACT

This paper gives an overview of our work on statistical machine translation of spoken dialogues, in particular in the framework of the VERBMOBIL project. The goal of the VERBMOBIL project is the translation of spoken dialogues in the domains of appointment scheduling and travel planning. Starting with the Bayes decision rule as in speech recognition, we show how the required probability distributions can be structured into three parts: the language model, the alignment model and the lexicon model. We describe the components of the system and report results on the VERBMOBIL task. The experience obtained in the VERBMOBIL project, in particular a large-scale end-to-end evaluation, showed that the statistical approach resulted in significantly lower error rates than three competing translation approaches: the sentence error rate was 29% in comparison with 52% to 62% for the other translation approaches. Finally, we discuss the integrated approach to speech translation as opposed to the serial approach as it is widely used nowadays.

1. INTRODUCTION

In comparison with written language, speech and especially spontaneous speech poses additional difficulties for the task of automatic translation. Typically, these difficulties are caused by errors of the recognition process, which is carried out before the translation process. As a result, the sentence to be translated is not necessarily well-formed from a syntactic point-of-view. Even without recognition errors, speech translation has to cope with a lack of conventional syntactic structures because the structures of spontaneous speech differ from that of written language.

The statistical approach shows the potential to tackle these problems for the following reasons. First, the statistical approach is able to avoid hard decisions at any level of the translation process. Second, for any source sentence, a translated sentence in the target language is guaranteed to be generated. In most cases, this will be hopefully a syntactically perfect sentence in the target language; but even if this is not the case, in most cases, the translated sentence will convey the meaning of the spoken sentence.

The organization of this paper is as follows:

- Section 2: **The Bayes Decision Rule For Written Language Translation.**
  We will present the Bayes decision rule and the resulting architecture for the translation of written language. A key component in this approach is the so-called alignment concept, which is similar to Hidden Markov models used in speech recognition and which will be considered in more detail.

- Section 3: **Experimental Results**.
  Although the methods presented apply both to *written* and *spoken* language, we will limit ourselves here to spoken language and report on the large-scale experiments that were carried out in the VERBMOBIL project.

- Section 4: **Speech Translation: The Integrated Approach.**
  As an alternative to the *serial* coupling of recognition and translation that is used in our and other systems as well, we will consider the *integrated* approach to recognition and translation and the corresponding form of the Bayes decision rule [11].

Whereas statistical modelling is widely used in speech recognition, there are so far only a few research groups that apply statistical modelling to the translation of written or spoken language. The presentation here is based on work carried out in the framework of the EUTRANS project [9] and the VERBMOBIL project [29].

2. BAYES DECISION RULE FOR WRITTEN LANGUAGE TRANSLATION

2.1. Principle

In machine translation for written language, the goal is the translation of a text given in a source language into a target language. We are given a source string $f_1^J = f_1 ... f_j ... f_J$, which is to be translated into a target string $e_1^I = e_1 ... e_i ... e_I$. For historical reasons [6], we use the symbols f (like French) for source words and the symbol e (like English) for target words. In this paper, the term *word* always refers to a *full-form* word. Among all possible target strings, we will choose the string with

The work reported here was partly carried out in the VERBMOBIL project (contract number 01 IV 701 T4) funded by the German Federal Ministry of Education, Science, Research and Technology and in the EUTRANS project (contract number IT-LTR-OS 30268) funded by the European Union.

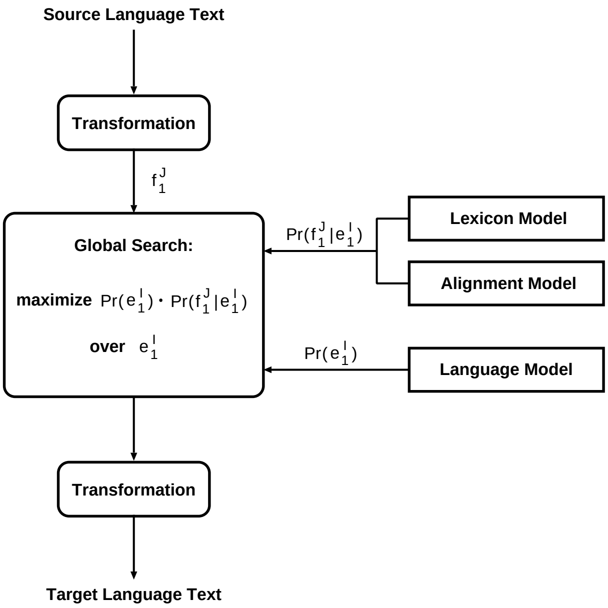

Figure 1: Architecture of the translation approach based on Bayes decision rule.

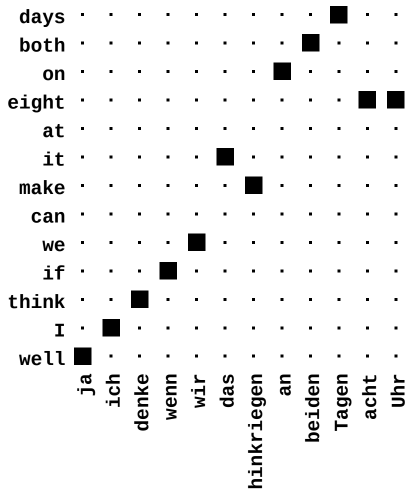

Figure 2: Word-to-word alignment.

the highest probability which is given by Bayes decision rule [6]:

$$\begin{aligned} \hat{e}_1^I &= \arg\max_{e_1^I} \{Pr(e_1^I|f_1^J)\} \\ &= \arg\max_{e_1^I} \{Pr(e_1^I) \cdot Pr(f_1^J|e_1^I)\} \quad . \end{aligned}$$

Here, $Pr(e_1^I)$ is the language model of the target language, and $Pr(f_1^J|e_1^I)$ is the string translation model which will be decomposed into lexicon and alignment models. The argmax operation denotes the search problem, i.e. the generation of the output sentence in the target language. The overall architecture of the statistical translation approach is summarized in Figure 1.

In general, as shown in this figure, there may be additional transformations to make the translation task simpler for the algorithm. The transformations may range from the categorization of single words and word groups to more complex preprocessing steps that require some parsing of the source string. We have to keep in mind that in the search procedure both the language and the translation model are applied *after* the text transformation steps. However, to keep the notation simple, we will not make this explicit distinction in the subsequent exposition.

2.2. Alignment Modelling

A key issue in modelling the string translation probability $Pr(f_1^J|e_1^I)$ is the question of how we define the correspondence between the words of the target sentence and the words of the source sentence. In typical cases, we can assume a sort of pairwise dependence by considering all word pairs (f_j, e_i) for a given sentence pair $(f_1^J; e_1^I)$. Here, we will further constrain this model by assigning each source word to *exactly one* target word. Later, this requirement will be relaxed. Models describing these types of dependencies are referred to as *alignment models* [6, 28].

When aligning the words in parallel texts, we typically observe a strong localization effect. Figure 2 illustrates this effect for the language pair German–English. In many cases, although not always, there is an additional property: over large portions of the source string, the alignment is monotone.

To arrive at a quantitative specification, we define the alignment mapping: $j \rightarrow i = a_j$, which assigns a word f_j in position j to a word e_i in position $i = a_j$. We rewrite the probability for the translation model by introducing the 'hidden' alignments $a_1^J := a_1...a_j...a_J$ for each sentence pair $(f_1^J; e_1^I)$. To structure this probability distribution, we factorize it over the positions in the source sentence and limit the alignment dependencies to a first-order dependence:

$$\begin{aligned} &Pr(f_1^J|e_1^I) = \\ &\quad = p(J|I) \cdot \sum_{a_1^J} \prod_{j=1}^{J} [p(a_j|a_{j-1}, I, J) \cdot p(f_j|e_{a_j})] \; . \end{aligned}$$

Here, we have the following probability distributions:

- the sentence length probability: $p(J|I)$, which is included here for completeness, but can be omitted without loss of performance;
- the lexicon probability: $p(f|e)$;
- the alignment probability: $p(a_j|a_{j-1}, I, J)$.

By making the alignment probability $p(a_j|a_{j-1}, I, J)$ dependent on the jump width $a_j - a_{j-1}$ instead of the absolute positions a_j, we obtain the so-called homogeneous hidden Markov model, for short HMM [28].

We can also use a *zero-order* model $p(a_j|j, I, J)$, where there is only a dependence on the *absolute* position index j of the source string. This is the so-called model IBM-2 [6]. Assuming a uniform alignment probability $p(a_j|j, I, J) = 1/I$, we arrive at the so-called model IBM-1.

These models can be extended to allow for source words having no counterpart in the translation. Formally, this is incorporated into the alignment models by adding a so-called 'empty word' at position $i = 0$ to the target sentence and aligning all source words without a direct translation to this empty word.

In [6], more refined alignment models are introduced by using the concept of fertility. The idea is that often a word in the target language may be aligned to several words in the source language. This is the so-called model IBM-3. Using, in addition, first-order alignment probabilities along the positions of the source string leads us to model IBM-4. Although these models take one-to-many alignments explicitly into account, the lexicon probabilities $p(f|e)$ are still based on single words in each of the two languages. In systematic experiments, it was found that the quality of the alignments determined from the bilingual training corpus has a direct effect on the translation quality [17]. By exchanging the role of target and source language in the training procedure, we found that the quality of the alignments could be significantly improved.

From a general point of view, the alignments can be interpreted as as a method for finding words or word groups that are equivalent in source language and target language. After these equivalences have been found, they may be modelled in various, data-driven approaches to build a translation system. Here, we will consider the so-called alignment templates (see next paragraph), but these equivalences may as well be used in finite-state transducers [7].

2.3. Alignment Templates

A general shortcoming of the baseline alignment models is that they are mainly designed to model the lexicon dependences between single words. Therefore, we extend the approach to handle word groups or phrases rather than single words as the basis for the alignment models [18]. In other words, a whole group of adjacent words in the source sentence may be aligned with a whole group of adjacent words in the target language. As a result, the context of words tends to be explicitly taken into account, and the differences in local word orders between source and target languages can be learned explicitly. Figure 3 shows some of the extracted alignment templates for a sentence pair from the VERBMOBIL training corpus. The training algorithm for the alignment templates extracts all phrase pairs which are aligned in the training corpus up to a maximum length of 7 words. To improve the generalization capability of the alignment templates, the templates are determined for bilingual word classes rather than words directly. These word classes are determined by an automatic clustering procedure [16].

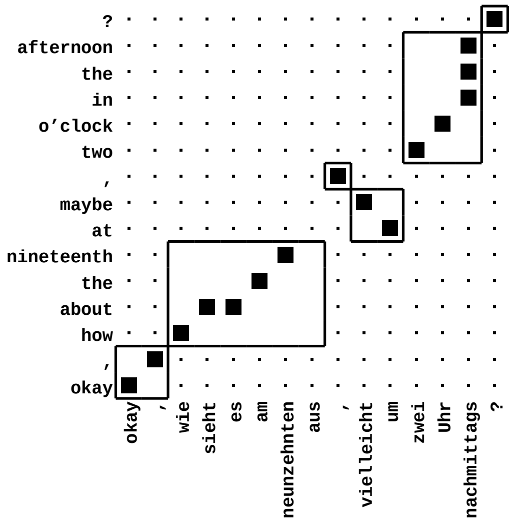

Figure 3: Example of a word alignment and of extracted alignment templates.

2.4. Search

The task of the search algorithm is to generate the most likely target sentence e_1^I of unknown length I for an observed source sentence f_1^J. The search must make use of all three knowledge sources as illustrated by Figure 4: the alignment model, the (bilingual) lexicon model and the language model. All three of them must contribute in the final decision about the words in the target language.

To illustrate the specific details of the search problem, we slightly change the definitions of the alignments:

- we use *inverted* alignments as in the model IBM-4 [6] which define a mapping from *target* to *source* positions rather the other way round.
- we allow *several* positions in the source language to be covered, i.e. we consider mappings B of the form:

$$B : i \rightarrow B_i \subset \{1, ..., j, ..., J\}$$

We replace the sum over all alignments by the best alignment, which is referred to as maximum approximation in speech recognition. Using a trigram language model $p(e_i|, e_{i-2}, e_{i-1})$, we obtain the following search criterion:

$$\max_{B_1^I, e_1^I} \prod_{i=1}^{I} \left[[p(e_i|e_{i-2}^{i-1}) \cdot p(B_i|B_{i-1}, I, J) \cdot \prod_{j \in B_i} p(f_j|e_i)] \right]$$

Considering this criterion, we can see that we can build up hypotheses of partial target sentences in a *bottom-to-top* strategy over the positions i of the target sentence e_1^i as illustrated in Figure 5. An important constraint for the alignment is that *all* positions of the source sentence should be covered exactly *once*. This constraint is similar to that of the travelling salesman problem where each city

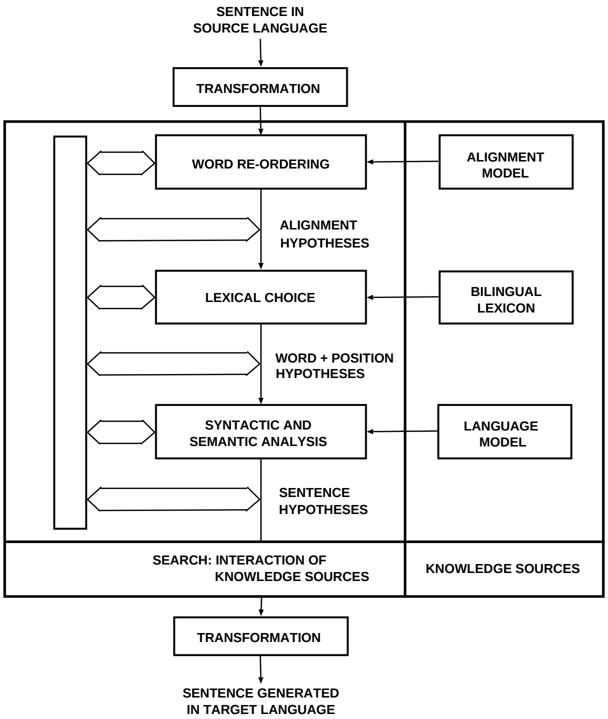

Figure 4: Illustration of search in statistical translation.

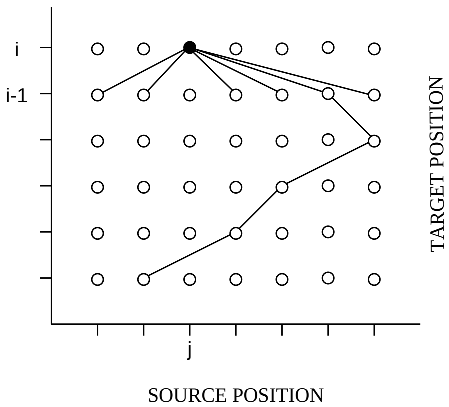

Figure 5: Illustration of bottom-to-top search.

has to be visited exactly once. Details on various search strategies can be found in [5, 12, 15, 19].

In order to take long context dependences into account, we use a class-based five-gram language model with backing-off. Beam search is used to handle the huge search space. To normalize the costs of partial hypotheses covering different parts of the input sentence, an (optimistic) estimation of the remaining cost is added to the current accumulated cost as follows. For each word in the source sentence, a lower bound on its translation cost is determined beforehand. Using this lower bound, it is possible to achieve an efficient estimation of the remaining cost.

3. EXPERIMENTAL RESULTS

3.1. The Task and the Corpus

Within the VERBMOBIL project [29], spoken dialogues were recorded. These dialogues were manually transcribed and later manually translated by VERBMOBIL partners (Hildesheim for Phase I and Tübingen for Phase II). Since different human translators were involved, there is great variability in the translations.

Each of these so-called dialogues turns may consist of several sentences spoken by the same speaker and is sometimes rather long. As a result, there is no one-to-one correspondence between source and target sentences. To achieve a one-to-one correspondence, the dialogue turns are split into shorter segments using punctuation marks as potential split points. Since the punctuation marks in source and target sentences are not necessarily identical, a dynamic programming approach is used to find the optimal segmentation points. The number of segments in the source sentence and in the test sentence can be different. The segmentation is scored using a word-based alignment model, and the segmentation with the best score is selected. This segmented corpus is the starting point for the training of translation and language models. Alignment models of increasing complexity are trained on this bilingual corpus [17].

A standard vocabulary had been defined for the various speech recognizers used in VERBMOBIL. However, not all words of this vocabulary were observed in the training corpus. Therefore, the translation vocabulary was extended semi-automatically by adding about 13 000 German–English word pairs from an online bilingual lexicon available on the web. The resulting lexicon contained not only word-word entries, but also multi-word translations, especially for the large number of German compound words. To counteract the sparseness of the training data, a couple of straightforward rule-based preprocessing steps were applied *before* any other type of processing:

- categorization of proper names for persons and cities,
- normalization of:
  – numbers,
  – time and date phrases,
  – spelling: `don’t` → `do not`,...
- splitting of
  German compound words.

Table 1 gives the characteristics of the training corpus and the lexicon. The 58 000 sentence pairs comprise about half a million running words for each language of the bilingual training corpus. The vocabulary size is the number of distinct full-form words seen in the training corpus. Punctuation marks are treated as regular words in the translation approach. Notice the large number of word singletons, i. e. words seen only once. The extended vocabulary is the vocabulary after adding the manual bilingual lexicon.

3.2. Offline Results

During the progress of the VERBMOBIL project, different variants of statistical translation were implemented, and experimental tests were performed for both text and speech input. To summarize these experimental tests, we briefly report experimental offline results for the following translation approaches:

- single-word based approach [24];
- alignment template approach [18];
- cascaded transducer approach [27]:
  unlike the other two-approaches, this approach requires a semi-automatic training procedure, in which the structure of the finite state transducers is designed manually. For more details, see [27].

The offline tests were performed on text input for the translation direction from German to English. The test set consisted of 251 sentences, which comprised 2197 words and 430 punctuation marks. The results are shown in Table 2. To judge and compare the quality of different translation approaches in offline tests, we typically use the following error measures [14]:

- mWER (multi-reference word error rate):
  For each test sentence in the source language, there are *several* reference translations in the target language. For each translation of the test sentence, the edit distances (number of substitutions, deletions and insertions as in speech recognition) to all reference translations are calculated, and the smallest distance is selected and used as error measure.
- SSER (subjective sentence error rate) [14]:
  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).

Both error measures are reported in Table 2. Although the experiments with the cascaded transducers [27] were not fully optimized yet, the preliminary results indicated that this semi-automatic approach does not generalize as well as the other two fully automatic approaches. Among these two, the alignment template approach was found to work consistently better across different test sets (and also tasks different from VERBMOBIL). Therefore, the alignment template approach was used in the final VERBMOBIL prototype system.

Table 1: Bilingual training corpus, recognition lexicon and translation lexicon (PM = punctuation mark).

		German	English
Training Text	Sentence Pairs	58 332	
	Words (+PMs)	519 523	549 921
	Vocabulary	7 940	4 673
	Singletons	44.8%	37.6%
Recognition	Vocabulary	10 157	6 871
Translation	Added Word Pairs	12 779	
	Vocabulary	11 501	6 867

Table 2: Comparison of three statistical translation approaches (test on text input: 251 sentences = 2197 words + 430 punctuation marks).

Translation Approach	mWER [%]	SSER [%]
Single-Word Based	38.2	35.7
Alignment Template	36.0	29.0
Cascaded Transducers	>40.0	>40.0

3.3. Integration into the VERBMOBIL Prototype System

The statistical approach to machine translation is embodied in the *stattrans* module which is integrated into the VERBMOBIL prototype system. We briefly review those aspects of it that are relevant for the statistical translation approach. The implementation supports the translation directions from German to English and from English to German. In regular processing mode, the *stattrans* module receives its input from the *repair* module [22]. At that time, the word lattices and best hypotheses from the speech recognition systems have already been prosodically annotated, i.e. information about prosodic segment boundaries, sentence mode and accentuated syllables are added to each edge in the word lattice [3]. The translation is performed on the single best sentence hypothesis of the recognizer.

The prosodic boundaries and the sentence mode information are utilized by the *stattrans* module as follows. If there is a major phrase boundary, a full stop or question mark is inserted into the word sequence, depending on the sentence mode as indicated by the *prosody* module. Additional commas are inserted for other types of segment boundaries. The *prosody* module calculates probabilities for segment boundaries, and thresholds are used to decide if the sentence marks are to be inserted. These thresholds have been selected in such a way that, on the average, for each dialogue turn, a good segmentation is obtained. The segment boundaries restrict possible word reordering between source and target language. This not only improves translation quality, but also restricts the search space and thereby speeds up the translation process.

3.4. Large-Scale End-to-End Evaluation

Whereas the offline tests reported above were important for the optimization and tuning of the system, the most important evaluation was the final evaluation of the VERBMOBIL prototype in spring 2000. This end-to-end evaluation of the VERBMOBIL system was performed at

Table 3: Sentence error rates of end-to-end evaluation (speech recognizer with WER=25%; corpus of 5069 and 4136 dialogue turns for translation German to English and English to German, respectively).

Translation Method	Error [%]
Semantic Transfer	62
Dialogue Act Based	60
Example Based	52
Statistical	29

the University of Hamburg [23]. In each session of this evaluation, two native speakers conducted a dialogue. They did not have any direct contact and could only interact by speaking and listening to the VERBMOBIL system.

Three other translation approaches had been integrated into the VERBMOBIL prototype system:

- a classical transfer approach [4, 8, 25], which is based on a manually designed analysis grammar, a set of transfer rules, and a generation grammar,
- a dialogue act based approach [20], which amounts to a sort of *slot filling* by classifying each sentence into one out of a small number of possible sentence patterns and filling in the slot values,
- an example-based approach [2], where a sort of nearest neighbour concept is applied to the set of bilingual training sentence pairs after suitable preprocessing.

In the final end-to-end evaluation, human evaluators judged the translation quality for each of the four translation results using the following criterion:

Is the sentence approximately correct: yes/no?

The evaluators were asked to pay particular attention to the semantic information (e.g. date and place of meeting, participants etc) contained in the translation. A missing translation as it may happen for the transfer approach or other approaches was counted as wrong translation. The evaluation was based on 5069 dialogue turns for the translation from German to English and on 4136 dialogue turns for the translation from English to German. The speech recognizers used had a word error rate of about 25%. The overall sentence error rates, i.e. resulting from recognition *and* translation, are summarized in Table 3. As we can see, the error rates for the statistical approach are smaller by a factor of about 2 in comparison with the other approaches.

In agreement with other evaluation experiments, these experiments show that the statistical modelling approach may be comparable to or better than the conventional rule-based approach. In particular, the statistical approach seems to have the advantage if robustness is important, e.g. when the input string is not grammatically correct or when it is corrupted by recognition errors.

Although both text and speech input are translated with good quality on the average by the statistical approach, there are examples where the syntactic structure of the produced sentence is not correct. Some of these syntactic errors are related to long range dependencies and syntactic structures that are not captured by the m-gram language model used. To cope with these problems, morpho-syntactic analysis [13] and grammar-based language models [21] are currently being studied.

4. SPEECH TRANSLATION: THE INTEGRATED APPROACH

In the Bayes decision rule, we have so far assumed written input, i.e. perfect input with no errors. When trying to derive a strict statistical decision rule for translation of spoken input, we are faced with the additional complication of speech recognition errors. So the question comes up of how to integrate the probabilities of the speech recognition process into the translation process. Although there have been activities in speech translation at several places [1, 10, 26], there has been not much work on this question of recognition/translation integration.

Considering the problem of speech input rather than text input for translation, we can distinguish three levels, namely the acoustic vectors $x_1^T = x_1...x_t...x_T$ over time $t = 1...T$, the source words f_1^J and the target words e_1^I:

$$x_1^T \rightarrow f_1^J \rightarrow e_1^I$$

From a strict point of view, the source words f_1^J are not of direct interest for the speech translation task. Mathematically, this is captured by introducing the possible source word strings f_1^J as hidden variables into the Bayes decision rule:

$$\begin{aligned}
&\arg\max_{e_1^I} Pr(e_1^I|x_1^T) \; = \\
&= \arg\max_{e_1^I} \left\{ Pr(e_1^I) \cdot Pr(x_1^T|e_1^I) \right\} \\
&= \arg\max_{e_1^I} \left\{ Pr(e_1^I) \cdot \sum_{f_1^J} Pr(f_1^J, x_1^T|e_1^I) \right\} \\
&= \arg\max_{e_1^I} \left\{ Pr(e_1^I) \cdot \sum_{f_1^J} Pr(f_1^J|e_1^I) \cdot Pr(x_1^T|f_1^J, e_1^I) \right\} \\
&= \arg\max_{e_1^I} \left\{ Pr(e_1^I) \cdot \sum_{f_1^J} Pr(f_1^J|e_1^I) \cdot Pr(x_1^T|f_1^J) \right\} \\
&\cong \arg\max_{e_1^I} \left\{ Pr(e_1^I) \cdot \max_{f_1^J} \left\{ Pr(f_1^J|e_1^I) \cdot Pr(x_1^T|f_1^J) \right\} \right\}
\end{aligned}$$

Here, we have made no special modelling assumption, apart from the reasonable assumption that

$$Pr(x_1^T|f_1^J, e_1^I) = Pr(x_1^T|f_1^J) \; ,$$

i. e. the target string e_1^I does not help to predict the acoustic vectors (in the source language) *if* the source string f_1^J is given. In addition, in the last equation, we have used the maximum approximation. Only in that special case of speech translation, at least from a strict point of view, there is the notion of a 'recognized' source word sequence f_1^J. However, this word sequence is very much determined by the combination of the language model $Pr(e_1^I)$ of the target language and the translation model $Pr(f_1^J|e_1^I)$. In contrast, in recognition, there would be only the language model $Pr(f_1^J)$.

When presenting the statistical approach to written language translation, the tacit assumption had been that the source sentence f_1^J was well formed. However, for speech input, this assumption is no more valid. Therefore, to take into account the requirement of 'well-formedness', we use a more complex translation model by including the dependence on the predecessor word:

$$p(f_j|f_{j-1}, e_{a_j}) \quad \text{in lieu of} \quad p(f_j|e_{a_j})$$

$$Pr(f_1^J|e_1^I) \;=\; \sum_{a_1^J} \prod_j \left[p(a_j|a_{j-1}, I) \cdot p(f_j|f_{j-1}, e_{a_j})\right]$$

For the sake of simplicity, here we have chosen the bigram dependence.

It is instructive to re-interpret already existing approaches for handling speech input in a translation task in the light of the Bayes decision rule for speech translation, even if these approaches are not based on stochastic modelling. The key issue in all these approaches is the question of how the requirement of having both a well-formed source sentence f_1^J and a well-formed target sentence e_1^I at the same time is satisfied. From the statistical point of view, this question is captured by finding suitable models for the *joint* probability $Pr(f_1^J, e_1^I) = Pr(e_1^I) \cdot Pr(f_1^J|e_1^I)$.

From the decision rule, it is clear that the translation process will have an effect on the recognition process only if the target language model $Pr(e_1^I)$ is sufficiently strong or, to be more exact, if its strength is comparable to that of the source language model $Pr(f_1^J)$. We mention the following approaches:

- In many systems, the method of n-best lists is used. The recognizer produces a list of n best source sentences, and the translation system works as a filter that selects one out of the n sentences using some suitable criterion. This joint generation and filtering process can be viewed as a crude approximation of the joint probability $Pr(f_1^J, e_1^I)$.

- When using finite-state methodology rather than a fully stochastic approach, the probability $Pr(f_1^J, e_1^I)$ is modelled by the finite-state network of the corresponding transducer, which is typically refined by domain and range restrictions [7, 26].

- In the extreme case, we might be only interested in the *meaning* of the target translation. Such an approach was used in [20] for the Verbmobil task. In Bayes decision rule, this case is captured by putting most emphasis on a *semantically* constrained language model $Pr(e_1^I)$.

However, it is clear that none of these approaches fully implements the integrated coupling of recognition and translation from a statistical point of view. We consider this integrated approach and its suitable implementation to be an open question for future research on spoken language translation.

5. SUMMARY

In this paper, we have given an overview of the statistical approach to machine translation and especially its implementation in the VERBMOBIL prototype system. The statistical system has been trained on about 500 000 running words from a bilingual German–English corpus. Translations are performed for both directions, i.e. from German to English and from English to German. Comparative evaluations with other translation approaches of the VERBMOBIL prototype system show that the statistical translation is superior, especially in the presence of speech input and ungrammatical input. In addition, we have presented the fully integrated approach to spoken language translation.

Acknowledgment

The author would like to thank the researchers in ITI Valencia and Aachen University of Technology who participated in the projects VERBMOBIL and EUTRANS and developed the approaches presented in this paper. In particular, the author would like to thank Enrique Vidal and Francisco Casacuberta for many discussions.

6. REFERENCES

[1] H. Alshawi, F. Xiang: English-to-Mandarin Speech Translation with Head Transducers. *Spoken Language Translation Workshop (SLT-97)*, Madrid, Spain, pp. 54-60, July 1997.

[2] M. Auerswald: Example-based machine translation with templates. In [29], pp. 418–427.

[3] A. Batliner, J. Buckow, H. Niemann, E. Nöth, V. Warnke: The prosody module. In [29], pp. 106–121.

[4] T. Becker, A. Kilger, P. Lopez, P. Poller: The Verbmobil generation component VM-GECO. In [29], pp. 481–496.

[5] A. L. Berger, P. F. Brown, J. Cocke, S. A. Della Pietra, V. J. Della Pietra, J. R. Gillett, J. D. Lafferty, R. L. Mercer, H. Printz,L. Ures: The Candide System for Machine Translation. *ARPA Human Language Technology Workshop*, Plainsboro, NJ, Morgan Kaufmann Publishers, pp. 152-157, San Mateo, CA, March 1994.

[6] P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, R. L. Mercer: The mathematics of statistical machine

translation: Parameter estimation. *Computational Linguistics*, Vol. 19, No. 2, pp. 263–311, 1993.

[7] F. Casacuberta: Finite-state transducers for speech input translation. *IEEE Automatic Speech Recognition and Understanding Workshop*, Madonna di Campiglio, Italy, Dec. 2001.

[8] M. C. Emele, M. Dorna, A. Lüdeling, H. Zinsmeister, C. Rohrer: Semantic-based transfer. In [29], pp. 359–376.

[9] EuTrans Project; Instituto Tecnológico de Informática (ITI, Spain), Fondazione Ugo Bordoni (FUB, Italy), RWTH Aachen, Lehrstuhl f. Informatik VI (Germany), Zeres GmbH Bochum (Germany): Example-Based Language Translation Systems. *Final report of the EuTrans project* (EU project number 30268), July 2000.

[10] A. Lavie, L. Levin, A. Waibel, D. Gates, M. Gavalda, L. Mayfield: JANUS: Multi-lingual translation of spontaneous speech in a limited domain. *2nd Conf. of the Association for Machine Translation in the Americas* pp. 252-255, Montreal, Quebec, Oct. 1995.

[11] H. Ney: Speech Translation: Coupling of Recognition and Translation. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing,* pp. I-517-520, Phoenix, AR, March 1999.

[12] H. Ney, S. Nießen, F. J. Och, H. Sawaf, C. Tillmann, S. Vogel: Algorithms for statistical translation of spoken language. *IEEE Trans. on Speech and Audio Processing* Vol. 8, No. 1, pp. 24–36, Jan. 2000.

[13] S. Nießen, H. Ney: Improving SMT quality with morpho-syntactic analysis. *18th Int. Conf. on Computational Linguistics*, pp. 1081-1085, Saarbrücken, Germany, July 2000.

[14] S. Nießen, F.-J. Och, G. Leusch, H. Ney: An evaluation tool for machine translation: Fast evaluation for MT research. *2nd Int. Conf. on Language Resources and Evaluation*, pp.39–45, Athens, Greece, May 2000.

[15] S. Nießen, S. Vogel, H. Ney, C. Tillmann: A DP based search algorithm for statistical machine translation. *COLING–ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th Int. Conf. on Computational Linguistics*, pp. 960–967, Montreal, Canada, Aug. 1998.

[16] F. J. Och: An efficient method to determine bilingual word classes. *9th Conf. of the European Chapter of the Association for Computational Linguistics*, pp. 71–76, Bergen, Norway, June 1999.

[17] F. J. Och, H. Ney: A comparison of alignment models for statistical machine translation. *18th Int. Conf. on Computational Linguistics*, pp. 1086-1090, Saarbrücken, Germany, July 2000.

[18] F. J. Och, C. Tillmann, H. Ney: Improved alignment models for statistical machine translation. *Joint SIGDAT Conf. on Empirical Methods in Natural Language Processing and Very Large Corpora*, pp. 20–28, University of Maryland, College Park, MD, June 1999.

[19] F. J. Och, N. Ueffing, H. Ney: An efficient A* search algorithm for statistical machine translation. *Data-Driven Machine Translation Workshop, 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 55–62, Toulouse, Juli 2001.

[20] N. Reithinger, R. Engel: Robust content extraction for translation and dialog processing. In [29], pp. 428–437.

[21] H. Sawaf, K. Schütz, H. Ney: On the use of grammar based language models for statistical machine translation. *6th Int. Workshop on Parsing Technologies*, pp. 231–241, Trento, Italy, Feb. 2000.

[22] J. Spilker, M. Klarner, G. Görz: Processing self-corrections in a speech-to-speech system. In [29], pp. 131–140.

[23] L. Tessiore, W. v. Hahn: Functional validation of a machine translation system: Verbmobil. In [29], pp. 611–631.

[24] C. Tillmann, H. Ney: Word re-ordering in a DP-based approach to statistical MT. *18th Int. Conf. on Computational Linguistics 2000,* Saarbrücken, Germany, pp. 850-856, Aug. 2000.

[25] H. Uszkoreit, D. Flickinger, W. Kasper, I. A. Sag: Deep linguistic analysis with HPSG. In [29], pp. 216–263.

[26] E. Vidal: Finite-state speech-to-speech translation. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Munich, Germany, pp. 111-114, April 1997.

[27] S. Vogel, H. Ney: Translation with Cascaded Finite-State Transducers. *ACL Conf. (Assoc. for Comput. Linguistics), Hongkong,* pp. 23-30, Oct. 2000.

[28] S. Vogel, H. Ney, C. Tillmann: HMM-based word alignment in statistical translation. *16th Int. Conf. on Computational Linguistics*, pp. 836–841, Copenhagen, Denmark, August 1996.

[29] W. Wahlster (Ed.): *Verbmobil: Foundations of speech-to-speech translations.* Springer-Verlag, Berlin, Germany, 2000.