A new search approach for interactive-predictive computer-assisted translation

Zeinab VAKIL Shahram KHADIVI
Human Language Technology Lab,
Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran
{Z.Vakil,Khadivi}@aut.ac.ir

ABSTRACT

Although significant improvements have been achieved in statistical machine translation (SMT), even the best machine translation technology is far from competing with human translators. An alternative approach to obtain high quality translation is to use a human translator who is assisted by an SMT. In interactive-predictive computer-assisted translation (IPCAT) paradigm, the human translator begins to type the translation of a given source text; by typing each character the MT system interactively offers the choices to complete the translation. Human translator may continue typing or accept the whole completion or part of it. In this paper, we propose a new search approach for increasing the performance of the IPCAT. This new search approach consists of a new search method and a hybrid back-off model. We achieve 2.3% and 1.16% absolute improvements by using the proposed search approach for two different corpora.

KEYWORDS : Statistical Machine Translation (SMT), Computer-Assisted Translation (CAT), Interactive-Predictive Computer-Assisted Translation (IPCAT), Prefix Search.

Proceedings of COLING 2012: Posters, pages 1261–1270,
COLING 2012, Mumbai, December 2012.

1 Introduction

Nowadays, with the expansion of global communications, the need for the translation has become a basic and important requirement, especially for international institutions and news agencies. Consider the following example to illustrate the importance of the translation in today world. In 2003, after the enlargement of the European Union, with a population of 453 million, the cost of the translation at all institutions, once translators are operating at full speed, was estimated at 807 M€ per year.

Recently, significant improvements have been achieved in statistical machine translation (MT), but still even the best machine translation technology is far from replacing or even competing with human translators. Because of the inability of existing MT systems for giving the correct and perfect translation, Researchers began to provide tools to facilitate and accelerate the translation process, instead of automatic translation. Already, Interactive computer-assisted translation systems are the latest version of these tools.

Interactive machine translation (IMT), first appeared as part of Kay's MIND system (Kay, 1973), where the user's role was to help with source-text disambiguation by answering questions about word sense, pronominal reference, prepositional-phrase attachment, etc. Later work on IMT, eg (Brown and Nirenburg, 1990; Maruyama and Watanabe, 1990; Whitelock et al., 1986), has followed in this vein, concentrating on improving the question/answer process by having less questions, more friendly ones, etc. Despite progress in these endeavors, the question/answer process remained in the systems of this sort. Finally these systems are only used where the cost of manually producing a translation is high enough to justify the extra effort. With introducing TransType project by (Foster et al., 1997), a major change in how the user interacts with the machine had occurred. In such an environment, human translators interact with a translation system that acts as an assistance tool and dynamically provides a list of translations (suffixes) which complete the part of the source sentence already translated (prefix). Also from 1997 to 2004, most of the given papers related to the various versions of the TransType project such as (Langlais et al., 2000 and 2002; Foster, 2002; Cubel et al., 2004).

In 2005, a new search strategy for giving suffix was proposed in (Bender et al., 2005). Also in (Barrachina et al., 2007), for creating search graph has been used finite state automata. Another important project in field of the interactive translation is Caitra project. Caitra is a web base project which is provided from an online platform and is based on the AJAX Web.2 technologies and the Moses decoder (Koehn, 2009a and 2009b). Another option which was added to the CAT is online learning; this option has been suggested in (Ortiz-Martínez et al., 2010). By this option, the interactive system can learn from user feedback and update itself statistical models.

In this paper, we will propose two new approaches to improve the performance of the interactive CAT system. To implement the interactive machine translation system, we use Moses as a statistical machine translation system. We extract of the Moses a search graph and offer a new search way of the graph which increases the quality of the suggestions of the interactive system. Also we offer a new back-off model which helps the system to suggest a suffix to the user in the some cases which the search graph does not consistent with the user prefix.

In the follow sections, the first we introduce the translation engine of our system. Next in the section three we describe interactive part of the system and our proposed approaches then we evaluate our system in section four.

2 Engine of translation

As mentioned in the introduction, we develop an interactive CAT for English to Germany by Moses system. Moses (Koehn et al., 2007) is a statistical machine translation system that allows us to automatically train translation models for English-Germany language pair. Indeed, Moses is translation engine of our interactive CAT. Also we use from Moses for offering a complementary translation to human translator. For giving a suitable suffix according to prefix, we created a graph by using hypotheses of Moses which are produced in decoding phase of the translation process of Moses. For better definition of the translation engine of our interactive CAT, we need to define statistical machine translation system and decoding phase of the Moses.

2.1 Statistical Machine Translation System

A statistical machine translation system allows us to automatically train translation models for any language pair by using parallel bilingual corpus and statistical theories. In statistical machine translation, we are given a source language sentence $F = f_1^J = f_1, \dots, f_j, \dots, f_J$, which is to be translated into a target language sentence $E = e_1^I = e_1, \dots, e_i, \dots, e_I$. Among all possible target language sentences, we will choose the sentence with the highest probability:

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

$$= \operatorname{argmax}\{Pr(e_1^I) \cdot Pr(f_1^J|e_1^I)\} \quad (2)$$

The decomposition into two knowledge sources in Equation 2 is known as the source channel approach to statistical machine translation (Brown et al., 1990). It allows an independent modelling of the target language model $Pr(e_1^I)$ and the translation model $Pr(f_1^J|e_1^I)$. The target language model describes the well-formedness of the target language sentence. The translation model links the source language sentence to the target language sentence. It can be further decomposed into the alignment and the lexicon models. The argmax operation denotes the search problem, i.e. the generation of the output sentence in the target language. We have to maximize over all possible target language sentences.

2.2 Decoding phase

The task of decoding in a machine translation system is to find the best scoring translation according to probabilistic scores of the language model and the translation model. This is a hard problem, since there are an exponential number of choices, given a specific input sentence. In fact, it has been shown that the decoding problem for the presented machine translation models is NP-complete (Knight, 1999; Udupa and Maji, 2006). In order to reduce the search space, we have to resort to a search heuristic. To this end, Moses organizes hypotheses into hypothesis stacks. If the stacks get too large, Moses prune out the worst hypotheses in the stack. One way to organize hypothesis stacks is based on the number of foreign words translated. One stack contains all hypotheses that have translated one foreign word; another stack contains all hypotheses that have translated two foreign words in their path, and so on.

3 Engine of Interaction

As described in the introduction, whenever user apply any change by keyboard in the translation, the system according to the modified translation, offers the completed translation. Now in this section, we want to investigate how the system is able to provide the completed translation based

on the prefix translation. For providing a completed translation, the system should seek the graph which is produced from hypotheses of the Moses decoder. As described in section 2-2, in decoding process of Moses, Hypotheses are organized in the stacks while we need to graph structure. Therefore the first task of the interactive component is to create a search graph from the Hypotheses into stacks of Moses. For creating the search graph, we reinstruct the organization of the hypotheses of the Moses from stacks to the graph by map data structure of C++. After finding the hypothesis which consistent with the prefix, the interactive component should give a completed translation to the user by using completed optimal path of that hypothesis in the search graph. In the next subsections, we will describe common search way and new our search way.

3.1 Edit Distance-Based Search

According to (Barrachina et al., 2007; Koehn, 2009a), for giving a completed translation to the user, we should find a node of the graph which has minimum edit distance with prefix; we call this approach, *edit distance-based search*. The purpose of the edit distance between two strings is the Levenshtein distance (Levenshtein, 1965) that defined as the minimum number of edits needed to transform one string into the other, with the allowable edit operations being insertion, deletion, or substitution of a single character.

This method is based on the assumption that a hypothesis which has minimum edit distance with prefix, has a greater chance to consistent with the desired translation of the user in the future than other hypotheses. If there are several hypotheses with minimum edit distance, we should compare cost of translation of the hypotheses together. The purpose of the cost of translation is summation of the current cost and the future cost of the hypothesis translation.

3.2 Using the translation cost in the search

The search method based on the edit distance has a fundamental inconsistency with the translation word graphs. The translation word graph has different hypotheses in terms of the orderings of the words (phrases); but not all the reordering possibilities due to the pruning that is applied during the generation of the word graph. Therefore, since the edit distance is only based on the deletion, insertion and substitution operations, this distance is not able to handle different ordering between the hypothesis and the reference sentences. I.e., we are only able to find those hypotheses which have similar ordering of words to the prefix of the user.

We explain this problem by using an example. We assume that the desired translation of the user is "**Newton is one of the greatest scientists who discovered gravity**" and our prefix is "**Newton is one of the greatest scientists w**". We also assume that only two translation hypotheses are available. The first hypothesis is "**one of the greatest scientists is Newton who discovered gravity**" which its translation cost is 0.0015. The second hypothesis is "**Newton gravity one of the greatest discovered which the greatest**" which its translation cost is 0.6812. In table 1, the edit distance between the first and the second hypotheses with the prefix is calculated. The numbers of this table are calculated according to Levenshtein algorithm (Levenshtein, 1965). Since the last word of the prefix is incomplete, we should find a complement to this prefix that its first word matches the last incomplete word of the prefix. According to the result of the Table 1 and the search method based on the edit distance, the second hypothesis is selected, while this hypothesis syntactically and semantically does not correct. The suffix which is offered according to the second hypothesis, is **"which the greatest"**. Obviously, this suffix is not compatible with the correct translation of the user.

	Hyp 1	one	of	the	greatest	scientists	is	Newton	who
Prefix	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Newton	**1**	1	2	3	4	5	6	6	7
Is	**2**	2	2	3	4	5	5	6	7
one	**3**	2	3	3	4	5	6	6	7
Of	**4**	3	2	3	4	5	6	7	7
the	**5**	4	3	2	3	4	5	6	7
greatest	**6**	5	4	3	2	3	4	5	6
scientists	**7**	6	5	4	3	2	3	4	5
w?	**8**	7	6	5	4	3	3	4	**4**

	Hyp 2	Newton	gravity	one	of	the	greatest	discovered	which
Prefix	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Newton	**1**	0	1	2	3	4	5	6	7
Is	**2**	1	1	2	3	4	5	6	7
one	**3**	2	2	1	2	3	4	5	6
Of	**4**	3	3	2	1	2	3	4	5
the	**5**	4	4	3	2	1	2	3	4
greatest	**6**	5	5	4	3	2	1	2	3
scientists	**7**	6	6	5	4	3	2	2	3
w?	**8**	7	7	6	5	4	3	3	**2**

TABLE 1 - The edit distance matrix between the first hypothesis and the user prefix.

According to the results obtained in the previous example, we can conclude that edit distance measure is not enough for finding a correct suffix. If we only emphasis on edit distance measure, our search may lead to find a hypothesis which the scores of the language and translation models is low; rationally, such hypothesis would not be acceptable in opinion of the user. To overcome this problem, we propose a new search approach. In this way, we use the weighted summation of the edit distance and the cost of the translation of the hypothesis in search process, that is:

$$compare\ measure\ = (\alpha \times D) + ((1 - \alpha) \times C) \quad (3)$$

Where D is edit distance between the hypothesis and prefix and C is the summation of the current and future translation cost of the hypothesis; this cost include both language and translation models.

The idea of this approach is stemmed from the reality that a translation hypothesis which its cost of translation and language models is lower than other hypotheses has more chance to be a correct translation and to be consistent with desired translation of the user in the future. We should note that, this search method might find hypotheses that do not match with the prefix at all due to the reordering of phrases like the previous example, and therefore we cannot generate a good offer to the user. However, we hope this method generates better offers to the user in overall. In the previous example, if we use new approach and set $\alpha = 0.2$, we will have:

$$Hyp1 = (0.2 \times 4) + (0.8 \times 0.0015) = 0.8,\ \ Hyp2 = (0.2 \times 2) + (0.8 \times 0.6812) = 0.96 \quad (4)$$

According to above result, the first hypothesis has lower cost than the second hypothesis, thus the first hypothesis will be selected.

The weights related to the edit distance and the cost of the translation, are empirically determined by development set of the bilingual corpus. Since we allow any amount of edit distance between the prefix and the word graph hypotheses and although we do not directly use the reordering of phrases, our IPCAT system is able to generate offers even there is not any hypotheses in the word graph with similar ordering to the user prefix. The results of the experiments are presented in Section 4. In the experiments, the weight of the edit distance and the translation cost are set to 0.2 and 0.8, respectively.

3.3 Back-Off models

In some cases, it is possible that any of the search method which are described in pervious sections, are not able to offer a suggestion to the user. This problem often occurs when the last word of the prefix is incomplete and there is not any phrase in the search graph that contains that partial word. This problem is solved in (Barrachina et al., 2007), by searching for a completion of the last word with the highest probability using only the language model. In this way isn't used any translation models, thus the degree of certainty of the suggestions which are produced by this way would be low. Now, we will propose a new approach which heightens the degree of certainty of the suggestions, but before explain it, we illustrate the problem of the pervious approach by an example.

Assume, we want to translate the Germany sentence **"het geluid van de muziek was luid"** to the English sentence "**the sound of music was loud**". Also we assume that the prefix is "**the sound of mu"** and there isn't any word in the search graph of the interactive system that contains the partial word of the prefix. If the interactive system only use language model, it will be possible that offers any word which starts with "mu" (such as **mu**sic, **mu**mmy, **mu**rmur, **mu**sketeer, **mu**tter, etc.), based on posterior probability of their occurrence after the penultimate word(s) of the prefix. According to the corpus which the language model has been trained it, each of the mentioned words can be selected. if the frequency of the phrase "sound of murmur" is more than others, then word **"murmur"** will be offered to the user; while if we attended to source sentence and translation model, we would select **"music"** word.

As we have explained, the selection process in above example was done only based on probability of the language model of the n-grams in the target language, without considering source sentence. In our proposed approach, we use IBM Model 1 (Brown et al., 1993) in addition to language model, to estimate the translation likelihood of the source sentence and candidate words (the purpose of candidate words is the words which start with the last partial word of the prefix). To achieve this goal, we use the weighted summation of the probability of the language model and the probability of the IBM-1 translation model.

Although using the IBM Model 1 in addition to the language model, has been proposed by (Ueffing and Ney, 2005), but its application is different from where we stand. They have used IBM Model-1 as a confidence measure for sub-sentences in the word graph, while we use the IBM Model-1 as a back-off model for words which are not available in the search graph.

IBM model 1 estimates the translation likelihood of a source language sentence $F = f_1^J = f_1 \dots f_j \dots f_j$, and a target language sentence $E = e_1^I = e_1 \dots e_i \dots e_I$, as:

$$Pr_{IBM-1}(E|F) = Pr\left(e_1^I \middle| f_1^J\right) = \prod_{i=1}^{I} \frac{1}{J+1} \sum_{j=1}^{J} p\left(e_i \middle| f_j\right) \quad (5)$$

According to equation 5, for obtaining the probability which a word e_i, be part of the translation of the source sentence f_1^J, we have:

$$Pr_{IBM-1}\left(e_i \middle| f_1^J\right) = \frac{1}{J+1} \sum_{j=1}^{J} p\left(e_i \middle| f_j\right) \quad (6)$$

In a hybrid model that has consisted of the both the language model and IBM-1 model, we have:

$$Pr_{IBM-1,LM}\left(e_i \middle| f_1^J\right) = \left(\alpha \times Pr_{LM}(e_i|e_{i-1})\right) + \left((1-\alpha) \times Pr_{IBM-1}\left(e_i \middle| f_1^J\right)\right) \quad (7)$$

Also we can use the higher IBM models such as IBM-2 or HMM instead of IBM model 1.

4 Evaluation of the purposed approach

For evaluating the performance of the interactive computer-assisted translation system, we need to estimate the effort of a human translator to produce the correct translations using the interactive system. To this end, the target translations which a real user would have in mind are simulated by the given reference(s). For each given source sentence, first the translation is produced by IPCAT system, then it is compared with a single reference translation to find the longest common character prefix. Afterwards, the first non-matching character is replaced by the corresponding reference character and then IPMT system offers a new complement to the given prefix. This process is iterated until a full match with the reference is obtained.

In order to evaluate the IPCAT system, we use KSR and KSMR metrics. The KSR is the number of key-strokes required to produce the single reference translation using the IPCAT system divided by the number of keystrokes needed to type the reference translation. The KSMR measure is the summation of KSR and MAR, which is the amount of all required actions either by keyboard or by mouse to generate the reference translation using the interactive machine translation system divided by the total number of reference characters.

We conduct the experiments on two different tasks: Xerox and Verbmobil. The Xerox is an English-German corpus, and the Verbmobil corpus is an English-Persian corpus, the Verbmobil corpus is originally an English-German corpus that we advanced it to an English-German-Persian corpus by translating a large part of English sentences to Persian. The statistics of these corpora are depicted in Table 2. The term OOVs in the table denotes the total number of occurrences of unknown words, the words which were not seen in the training corpus.

		Xerox		**Verbmobil**	
		English	Germany	English	Persian
Train	Sentences	47 619		22 642	
	Running words	528 779	467 633	254 665	233 948
	Vocabulary size	9 816	16 716	2 696	5 405
	Singletons	2 302	6 064	1 016	2 501
Dev	Sentences	700		276	
	Running words	8 823	8 050	5358	3 339
	OOVs	56	108	198	200
Eval	Sentences	862		250	
	Running words	10 019	10 094	2 871	2 692
	OOVs	58	100	142	193

TABLE 2 - The statistics of the Xerox and Verbmobil corpora.

4.1 Evaluation of the experiment result

In the first experiment, we evaluate the proposed search method which described in section 3-2. This method is based on the weighted summation of the edit distance and the cost of generating the complement translation for a given prefix in the word graph. In contrast, the previous method is only based on the edit distance measure. The results of the experiments are shown in table 3. According to the results, the proposed method is superior to the previous method and both the KSR and the KSMR measures are decreased. Therefore, we could conclude using the hypotheses

which have the lower cost in terms of the language and translation models in addition to the edit distance with a given prefix, lead to improve the results of the IPCAT systems.

The second experiment is conducted to evaluate the proposed back-off model. The new back-off model is a hybrid model which consists of IBM-1 and language models. The experimental results are shown in table 3, in the rows where the back-off models set to 'No'. As we expected, the proposed back-off model obtained better results than the previous back-off model, which is purely based on the language model. This improvement is due to the use of two knowledge sources namely source sentence and target language to estimate the back-off model, instead of just using the target language. Obviously with more information, our system gives better suffix to the user. Although, the result of the hybrid back-off model has been better than language model, but the difference between the results of these models is very small. The cause of this small difference may be that the desired translation of the user has the words which are not available in the training corpus. In such cases, neither language model nor IBM-1 model could suggest any suffix to the user.

Also we used IBM-2 model instead of IBM-1, but unfortunately, we it does not lead to obtain a better result, the reason of this result may be that the IBM-2 model apply more restriction than IBM-1 model.

	Back-off model	Xerox En→De		Verbmobil En→Pe	
		KSR	KSMR	KSR	KSMR
Edit distance	No	20.46	28.57	29.27	40.09
	IBM-1	16.13	25.43	25.47	37.66
	LM	15.25	24.31	24.39	36.68
	IBM-1 + LM	15.27	24.31	24.18	36.47
Edit distance + Translation cost	No	19.10	26.27	28.58	38.93
	IBM-1	14.46	22.67	24.64	36.35
	LM	13.88	22.00	23.59	35.46
	IBM-1 + LM	13.87	21.97	23.31	35.14

TABLE 3 - The results of various types of back-off models and search methods.

5 Conclusion

The goal of this paper was to develop an interactive computer assisted translation system. We recognized the defect of the edit distance measurement and offered new search way based on a combined measurement which consisted of edit distance and cost of translation. Edit distance measure does not consider reordering of phrase; thus by using this measure, two sentences "**Newton is one of the greatest scientists**" and "**one of the greatest scientists is Newton**" would have four edit distance. While by considering the reordering operation, the edit distance between these sentences would be only two. In this paper we didn't insert the reordering of the phrase operation, but we tried to decrease the defect of the edit distance measure by considering translation cost. We could achieve 2.3% and 1.16% improvements by using our offered measure search in Xerox and Verbmobil corpora respectively. Also we obtained 0.3% improvement by using new back-off model in the Verbmobil corpus.

Reference

Barrachina, S., Bender, O., Casacuberta, F., Civera, J., Cubel, E., Khadivi, S., Lagarda, A., Ney, H., Tomás, J., Vidal, E. and Vilar, J. M. (2007). *Statistical Approaches to Computer-Assisted Translation*, Computational Linguistics, Volume 35, pp. 3-28.

Bender, O., Hasan, S., Vilar, D., Zens, R. and Ney, H. (2005). *Comparison of generation strategies for interactive machine translation*, In Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT 05), pp. 33–40.

Brown, R.D., Nirenburg, S. (1990). *Human-computer interaction for semantic disambiguation,* In Processing of the International Conference on Computational Linguistics (COLING), PP. 42-47.

Brown, P.F., Cocke, J., Della Pietra, S.A., Della Pietra, V.J., Jelinek, F., Lafferty, J.D., Mercer, R.L., Roossin, P.S. (1990). *A Statistical Approach to Machine Translation*, Computational Linguistics, Vol. 16, No. 2, pp. 79–85.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J. and Mercer, R. L. (1993). *The mathematics of statistical machine translation: Parameter estimation*, Computational Linguistics, pp. 263–311.

Cubel, E., González, J., Lagarda, A. L., Casacuberta, F., Juan, A. and Vidal, E. (2004). *Adapting finite-state translation to the TransType2 project,* Proceedings of the Joint Conference combining the 8th International Workshop of the European Association for Machine Translation.

Foster, G., Isabelle, P. and Plamondon, P. (1997). *Target-Text Mediated Interactive Machine translation*, in Kluwer Academic Publishers, pp. 175–194.

Foster, G. (2002). *Text Prediction for Translators,* Ph.D. thesis, Universit´e de Montr´eal, Canada.

Kay, M. (1973). *The MIND system,* in Natural Language Processing, pp. 155-188.

Knight, K. (1999). *Decoding complexity in word replacement translation models*, Computational Linguistics, 25(4):607–615.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C. J., Bojar, O., Constantin, A. and Herbst, E. (2007). *Moses: Open source toolkit for statistical machine translation*, In ACL Demo and Poster Session, Available: http://www.statmt.org/moses/.

Koehn, P. (2009a). *A Process Study of Computed Aided Translation*, Kluwer Academic Publishers.

Koehn, P. (2009b). *A web-based interactive computer aided translation tool*, In Proceedings of the ACL Interactive Poster and Demonstration Sessions.

Levenshtein, V. I. (1965). *Binary Codes Capable of Correcting Deletions, Insertions, and Reversals*. Soviet Physics - Doklady, Vol. 10 No. 8 pp. 707-710.

Langlais, P., Foster, G., and Lapalme, G. (2000). *TransType: a computer-aided translation typing system*, In Proceedings of the NAACL/ANLP Workshop on Embedded Machine Translation Systems, pp. 46–52.

Langlais, P., Lapalme G. and Loranger, M. (2002). *TRANSTYPE: Development–Evaluation Cycles to Boost Translator's Productivity*, in Kluwer Academic Publishers, pp. 77–98.

Maruyama, H., Watanabe, H. (1990). *An interactive Japanese parser for machine translation,* In Processing of the International Conference on Computational Linguistics (COLING), pp. 257-262.

Ortiz-Martínez, D., García-Varea, I. and Casacuberta, F. (2010). *Online Learning for Interactive Statistical Machine Translation*, In The 2010 Annual Conference of the North American Chapter of the ACL, pp. 546–554.

Tillmann, C. (2001). *Word re-ordering and dynamic programming based search algorithms for statistical machine translation*, PhDthesis, Computer Science Department, RWTH Aachen, Germany.

Ueffing, N. and Ney, H. (2005). *Application of Word-Level Confidence Measures in Interactive Statistical Machine Translation*, In Proceedings of EAMT 2005 (10th Annual Conference of the European Association for Machine Translation), pp. 262-270, Budapest, Hungary.

Udupa, U. and Maji, H. K. (2006). *Computational Complexity of Statistical Machine Translation*, 11st Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference, Italy.

Whitelock, P. J., McGee Wood, M., Chandler, B. J., Holden, N. and Horsfall, H. J. (1986). *Strategies for interactive machine translation: the experience and implications of the UMIST Japanese project*, In Proceedings of the International Conference on Computational Linguistics (COLING), pages 329-334.

Zens, R., Och, F. J. and Ney, H. (2002). *Phrase-Based Statistical Machine Translation*, in Springer-Verlag Berlin Heidelberg, pp. 18–32.