

Example-Based Machine Translation Using a Dictionary of Word Pairs

Reinhard Rapp, Carlos Martin Vide

Universitat Rovira i Virgili
Research Group on Mathematical Linguistics
Plaza Imperial Tarraco, 1. 43005 Tarragona. Spain
reinhardrapp@gmx.de carlos.martin@urv.net

Abstract

Machine translation systems, whether rule-based, example-based, or statistical, all rely on dictionaries that are in essence mappings between individual words of the source and the target language. Criteria for the disambiguation of ambiguous words and for differences in word order between the two languages are not accounted for in the lexicon. Instead, these important issues are dealt with in the translation engines. Because the engines tend to be compact and (even with data-oriented approaches) do not fully reflect the complexity of the problem, this approach generally does not account for the more fine grained facets of word behavior. This leads to wrong generalizations and, as a consequence, translation quality tends to be poor. In this paper we suggest to approach this problem by using a new type of lexicon that is not based on individual words but on pairs of words. For each pair of consecutive words in the source language the lexicon lists the possible translations in the target language together with information on order and distance of the target words. The process of machine translation is then seen as a combinatorial problem: For all word pairs in a source sentence all possible translations are retrieved from the lexicon and then those translations are discarded that lead to contradictions when constructing the target sentence. This process implicitly leads to word sense disambiguation and to language specific reordering of words.

1. Introduction

According to Hutchins (1987), research on Machine Translation (MT) combines an intellectual challenge, a worthy motive, and an eminently practical objective. The challenge is to produce high quality translations, the motive to remove language barriers which hinder global communication and international understanding. And the practical objective is the development of economically viable systems to satisfy a growing demand for translations in the globalized world.

However, up to now despite considerable efforts (e.g. large scale projects such as EUROTRA and Verbmobil) a breakthrough could not be achieved, and some researchers (e.g. Kay, 1995) came to the conclusion that the idea of a fully automatic high-quality translation of arbitrary texts is not a realistic goal for the foreseeable future. A decade ago, Kay's analysis was that "Research in machine translation has developed traditional patterns which will clearly have to be broken if any real progress is to be made" (Kay, 1995). We believe that this break with traditional patterns may have happened in the meantime. In our view, the recent statistical and machine learning approaches bring up a new quality which is in better accordance with psycholinguistic principles (e.g. the law of association by contiguity) and has the potential to eventually succeed.

But let us look back at the history of machine translation. For several decades the *rule-based approach* has been dominating (Slocum, 1989; Hutchins & Somers, 1992; Arnold et al., 1994). It is based on the concept of grammatical well-formedness which is to be achieved by an appropriate rule base. The main advantage of the rule-based approach is that the rules tend to be compact and are tailored towards being intellectually comprehensible by linguists.

In contrast, with the *statistical approach* (pioneered by Brown et al, 1990) the distinction between grammatically well-formed and ill-formed sentences is considered to be fluent. Linguistic rules are replaced by statistical processes such as Hidden Markov Models (Rabiner, 1990), or translation is seen as a decoding problem related to cryp-

tography (Weaver, 1949). The parameters required for the statistical models are automatically extracted from the distribution of words in large text corpora. To put it simple, a sentence is translated by looking up each word in a bilingual dictionary and then constructing all possible sequences of the retrieved words. From the many candidate translations obtained in this way, the best is selected by determining the occurrence probability of each of its subsequences on the basis of a large corpus in the target language, and by assigning each candidate an overall probability. A problem with this approach is that only short subsequences of typically two or three words can be considered, as otherwise the probability estimates tend to be poor due to the problem of data sparseness. As a consequence, the approach can not take long distance dependencies into account, and translation quality is limited.

Like the statistical approach, the *example-based approach* (sometimes referred to as *translation by analogy*) is also data-oriented (Sato & Nagao, 1990). It tries to utilize human translation know how by using a database of previously translated texts. If a new sentence is to be translated, its structural similarity to some relevant source sentences in the database is computed, and a target sentence is constructed by using appropriate parts of the relevant source sentences. For example, if the translations "The man rides the car" and "John goes to the bank" are known, it may be possible to draw conclusions for the translation of the sentence "The man goes to the bank".

Let us mention that the distinction between the three paradigms is in practice not as clear cut as described above. For example, rule-based systems may be enhanced by statistical heuristics to reduce the search space. Statistical systems usually not only use a language model based on the target language, but also a so-called translation model derived from parallel texts. In essence, the translation model is an automatically extracted bilingual dictionary that includes translation probabilities for each pair of a source and a target language word. And example-based machine translation may well be implemented as a mix of

rule-based and statistical methods. Table 1 shows a comparison of the three approaches.

The current situation in the competition of the three paradigms is that the major commercial machine translation systems are still rule-based, but that over the last decade statistical and example-based systems have been rapidly gaining ground. Their main advantage is that they can be quickly adapted to new language pairs since the algorithms are almost language independent and most language specific information is automatically derived from parallel corpora.

Conceptually, the methodology that we describe in the next section (which is referred to as the *combinatorial approach*) can be seen as a mixture of the statistical and the example-based approach. Like these it has the advantage of being a self-learning system that – provided there is enough data in the form of previously translated texts – easily adapts to new language pairs. However, it is novel in that it relies on a new type of dictionary with a potentially far higher information content than conventional ones, and that it uses a new type of translation engine.

	R	S	E
Large parallel corpora are required.	+	-	-
The strength of developers lies in abstract language modeling.	+	-	-
Most of the knowledge required for MT is of inductive rather than deductive nature. (Linguistics is derived from natural languages, not vice versa.) Inductive knowledge can be extracted automatically and efficiently from corpora.	-	+	+
Today's computers are capable of processing large quantities of texts.	-	+	+
Vague intuitions can be modeled more easily statistically than by explicit rules.	-	+	+
Ensuring the consistency and integrity of a large rule-base is difficult, especially if its compilation takes a long time and if several people are involved.	-	+	+
The automatic extraction of grammatical rules from corpora is an unsolved problem.	-	+	+
Local adjustments to the rule base do not ensure a global improvement of translation quality. Over-adaptation to a specific text (sort) can easily happen.	-	+	+
Especially in speech many utterances are ungrammatical. Therefore, robustness is desirable.	-	+	+
The many kinds of statistical dependencies in texts require the computation of so many parameters that in practice only short distance dependencies can be taken into account.	+	-	+
Adaptation to new domains is performed automatically.	-	+	+

Table 1: Comparison of the rule-based (R), the statistical (S), and the example-based (E) approach in machine translation (inspired by Su & Chang, 1992). The symbols + and - indicate if the respective criterion is to be seen as an advantage or a disadvantage.

2. Approach

In our analysis of the current state of the art in machine translation we come to the conclusion that the bilingual dictionaries that are the basis for translation in current approaches contain too little information and, as a consequence, many decisions made by the translation engines are based on guesses instead of exact knowledge. The degree in which this statement is true varies of course between different systems, as for example in statistical machine translation the translation model and the language model are intended to make up for such deficits. Nevertheless the limited translation quality that is achieved in practice indicates that there is room for improvement.

The idea behind the combinatorial approach proposed here is that conventional dictionaries that are in essence mappings between individual words of the source and the target language are not appropriate as a basis for machine translation as their information content is far from adequate. In most cases they contain no information on word reordering and word sense disambiguation, two of the most important issues in machine translation. Currently, these issues are dealt with in the translation engines, but (even with data-oriented approaches) these tend to be compact and relatively unspecific with regard to the more fine grained properties of the individual word. This leads to errors and to overgeneralization and as a consequence translation quality tends to be poor.

To better account for the properties of the individual word, we therefore suggest to include information on word reordering and word sense disambiguation in the lexicon, thereby considerably expanding its information content. Our proposal is to use a new type of lexicon which is not based on individual words but on pairs of words. For each pair of words in the source language the lexicon lists the possible translations in the target language together with information on the order and distance of the target words. The process of machine translation is then seen as a combinatorial problem: For all word pairs in a source sentence all possible translations are retrieved from the lexicon and then those translations are discarded that lead to contradictions when constructing the target sentence. This process implicitly leads to word sense disambiguation and to language specific reordering of words. In comparison to state-of-the-art statistical systems, the method has the advantage that long distance dependencies between words can be taken into account.

As there are many more word pairs than individual words in a language, the information content of our lexicon is considerably higher than that of any conventional lexicon. On one hand, this gives the potential for better translations. On the other hand, it is not realistically possible to construct such a lexicon manually. For this reason, part of our work deals with a method on how to derive such a lexicon automatically from previously translated texts.

Let us illustrate our approach by giving an example for the language pair German → English. Note that for clarity this example is simplified and does not fully reflect the complexity of the problem. Our starting point is a lexicon of word pairs that is assumed to be given. Some sample entries of our lexicon could, for example, look as shown in the following table:

SOURCE LANGUAGE	TARGET LANGUAGE
dann kauft	then * buys
ein Rennrad	a racing bike a racing bicycle
er ein	he * a he * an
kauft er	he buys

In our example we assume that the entries in the source language only comprise word pairs that occur as direct neighbors in a large parallel corpus. In the column for the target language all translations as observed in this corpus are listed. The wildcard * indicates that the words in the target language are not direct neighbors but are separated by one or several other words. This reflects different word orders in the source and the target language.

To translate a sentence, its word pairs are looked up in the dictionary. In case of multiple translations those are preferred that lead to maximum mutual support which can be measured as phrasal overlap. If, for example, the German sentence “*dann kauft er ein Rennrad*”¹ is to be translated to English, in the first step the translation of “*dann kauft*” is looked up in the dictionary as “*then * buys*”. Looking up the next word pair (“*kauft er*” → “*he buys*”) gives evidence that – in order to maximize the overlap – the wildcard in “*then * buys*” is to be replaced by the personal pronoun “*he*”. Thus the translation of “*dann kauft er*” is “*then he buys*”.

For the translation of the next word pair, namely “*er ein*” the dictionary lists two possibilities: “*he * a*” and “*he * an*”. Both show the same amount of overlap with the first three words of the translation. The decision which one to choose can only be made after looking up the next word pair, namely “*ein Rennrad*”. Here we only get an overlap if “*he * a*” is chosen. The two possible translations of “*ein Rennrad*”, namely “*a racing bike*” or “*a racing bicycle*” both show the same overlap, which indicates that there is no preference for any of them. This means that for the translation of the full sentence we obtain two correct solutions:

then he buys a racing bike
then he buys a racing bicycle

Although in the above example for clarity we have described the translation process as being sequential, a better algorithm does not simply proceed from left to right but instead in a process of mutual reinforcement aims to find the global optimum of all possible combinations. Note that as we always consider pairs of words we obtain an implicit disambiguation effect.² This effect can be enhanced if we not only consider direct neighbors but all pairs of words that can be extracted from the source sentence. Also, as omissions and contradictions resulting from the lookup process can not be ruled out, a weighting scheme is required that gives each item retrieved from the dictionary a translation probability.

¹ A word-by-word translation of this sentence is “Then buys he a racing bicycle”.

² As experience tells, in most cases, by looking at context words ambiguities can be resolved.

Designing and realizing an appropriate algorithm requires considerable effort, which is why this core part of the system is still in an experimental stage. Possible solutions include statistical and machine learning approaches, algebraic methods, and artificial neural networks (e.g. backpropagation networks, self organizing maps, constraint satisfaction networks, or Boltzman machines).

An alternative solution is based on a dimensionality reduction using the algebraic method of singular value decomposition. That is, the translations found in our dictionary of word pairs are seen as spanning a multidimensional semantic space. By reducing the dimensionality of this space a generalization effect is achieved and possible contradictions are removed in an optimal way. This process can be compared to finding the global optimum using an artificial neural network. However, the advantage of the singular value decomposition is that it can be guaranteed to always find the global optimum, whereas neural networks have a tendency to get stuck in local optima. Due to this advantage, previous research on thesaurus construction and word sense induction based on such principles has been promising (Rapp, 2004a and 2004b).

For the translation algorithm described above a dictionary of word pairs is required. We created such a dictionary by performing a word alignment on a previously translated corpus, and by retrieving the dictionary from this resource. However, the process of word alignment is difficult as especially for function words alignments can be ambiguous or even impossible. Fortunately, we could draw on a lot of previous work on word alignment which made the task easier.

3. Procedure

The work reported here is part of an ongoing larger project whose overall design is shown in Table 2. As our parallel text collection we use the English/German part of the Proceedings of the European Parliament (Europarl corpus v2, years 1996 to 2003; Koehn, 2002) which is available on the world wide web in sentence-aligned form at <http://people.csail.mit.edu/koehn/publications/europarl/>. Documentation on an earlier corpus of parliamentary proceedings (MLCC-corpus), which is available from the European Language Resource Association (ELRA), can be found in Armstrong et al. (1998).

A version of the Europarl corpus that had been converted to lowercase letters has been word aligned for us by Jörg Tiedemann (2003) using the GIZA++ program which is freely available for research purposes (Och & Ney, 2003; <http://www.fjoch.com/GIZA++.html>). As the corpus has a size of about 25 million words per language, and as GIZA++ has higher memory requirements than were available at the time, to conduct the word alignment the corpus was split into 20 segments of about equal size. It can be assumed that splitting the corpus has a somewhat negative effect on the quality of the alignments.

Let us mention that it is our impression that a breakthrough in alignment issues has not yet been achieved, as the results on sentence alignment for the Europarl-corpus as currently published on the web are based on the pioneering Gale & Church algorithm (Gale & Church, 1993), and as sentence and word alignment is still a very active research area (Deng, Kumar & Byrne, in print).

Given the GIZA++ output as shown in Table 3, the extraction of the dictionary of word pairs has been per-

formed using a program that looks up the translations for each pair of consecutive words, as exemplified in Figure 1. It turns out that the resulting dictionary contains many errors, as each of the numerous misalignments results in an erroneous entry.

Assuming that most misalignments should be unique, in an attempt to reduce the number of errors we eliminated all dictionary entries that were based on alignments that occurred only once in the corpus.³ This reduced the number of dictionary entries from 8,698,054 to 1,381,892. To give an impression of the effects of this heuristic, Table 4 shows the corpus frequencies of a few sample alignments. Note that although we have selected familiar terms, the frequencies are mostly in the lower 1-digit range, which illustrates that the problem of data sparseness is severe for word pairs.


Acquisition of parallel corpora	
Acquisition of non-parallel corpora	
Conduct word alignment	
Generate dictionary of word pairs	
Generate conventional dictionary	from parallel texts
	from comparable texts
Design and implement translation engine	
Adopt algorithm for automatic evaluation	
	Translate sample texts
	Evaluate automatic translations
	Improve translation engine

Table 2: Steps of the translation project.

Sentence pair (1) source length 3 target length 4 alignment score : 0.000459411
resumption of the session
NULL ({ }) wiederaufnahme ({ 1 2 }) der ({ 3 }) sitzungsperiode ({ 4 })
Sentence pair (2) source length 33 target length 40 alignment score : 3.7275e-65
i declare resumed the session of the european parliament adjourned on friday 17 december 1999 , and i would like once again to wish you a happy new year in the hope that you enjoyed a pleasant festive period .
NULL ({ 6 17 23 30 31 }) ich ({ 1 }) erkläre ({ 2 }) die ({ 4 }) am ({ 11 }) freitag ({ 12 }) , ({ }) dem ({ }) 17. ({ 13 }) dezember ({ 14 }) unterbrochene ({ 10 15 }) sitzungsperiode ({ 5 }) des ({ 7 }) europäischn ({ 8 }) parlaments ({ 9 }) für ({ }) wiederaufgenommen ({ 3 }) , ({ 16 }) wünsche ({ 18 19 20 24 }) ihnen ({ 25 }) nochmals ({ 21 22 }) alles ({ }) gute ({ }) zum ({ }) jahreswechsel ({ 26 27 28 29 35 36 37 38 39 }) und ({ }) hoffe ({ 32 }) , ({ }) daß ({ 33 }) sie ({ 34 }) schöne ({ }) ferien ({ }) hatten ({ }) . ({ 40 })

Table 3: GIZA++ output: first two sentences of the Europarl corpus. The numbers in brackets refer to the word numbers in the English sentences. English words for which no correspondences could be identified are assigned to the NULL-item.

³ This does not account for undesired effects resulting from multiple occurrences of identical sentence pairs.

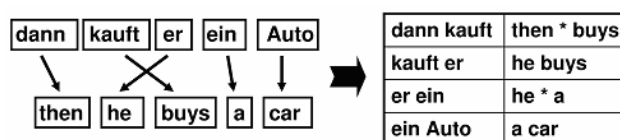


Fig. 1: Creating the dictionary of word-pairs from a word-aligned corpus.

beträchtliche zahl (considerable number)	10	considerable number
	1	huge numbers
	2	many
	1	number
	1	proposed a number
	1	quite * number
	5	significant number
	1	significant numbers
	4	substantial number
	1	viewed many
hohe verantwortung (high responsibility)	1	great responsibility
	1	have much * answer
	2	heavy responsibility
	1	high
	1	high degree * responsibility
	1	major responsibility
	1	responsibilities
nicht vertraut (not familiar)	5	responsibility
	1	in
	1	lack * skill
	1	never been familiar with
	1	not
	6	not familiar
	1	not familiar with them
1	they * not	
1	will	

Table 4: Corpus frequencies of a few sample word pair alignments. Only translations of frequency 2 and higher (printed in bold) were retained in the dictionary.

As a consequence, we cannot expect that the dictionary contains all word pairs that we require for the translation of new sentences. Therefore, as a back off strategy in addition to the dictionary of word pairs a conventional dictionary based on individual words is needed. Such a dictionary can also be derived from the word-aligned corpus. However, to get maximum coverage it is desirable to expand this dictionary using algorithms that are capable of identifying word translations from comparable corpora. The comparable corpora can, for example, be newspaper texts which are available in large quantities and therefore have the potential to improve coverage significantly. In a previous paper (Rapp, 1999) we described an associative algorithm that does this with good success, and the work of Koehn & Knight (2002) showed that for related languages orthographic clues (cognates) can also be useful. We refer to this literature and the work of Pascale Fung (e.g. Fung & Yee, 1998) for further details.

The core part of the system is the translation engine which realizes the combinatorial approach. At the present stage a brute force method has been implemented that generates all possible combinations of the translation sequences found in the word pair dictionary. It then chooses the combination where the sequences show maximum

overlap, i.e. where the number of words that occur in several sequences is highest. However, as the number of combinations increases exponentially with the number of words, this method is only suitable for short sentences. It may be useful to test the potential of the method, but for a practical system a more efficient implementation (possibly along the lines suggested in section 2) is mandatory.

As it is of critical importance for the quality of the translation output, the translation engine needs to be optimized. As indicated in Table 2, the intention is to do this in several modify-evaluate cycles. That is, after each modification of the algorithm the output will be evaluated and further modifications will be made in accordance with the results of the evaluation. The metric to be adopted for evaluation will be the BLEU-score (Papineni et al., 2002) which is readily available and well suited for this purpose as it allows to automatically compare the agreement of the machine translation with a given human translation.

4. Results

Tables 5 and 6 show the results for the translation of two German sentences into English. The first sentence is “*ich kann mich nicht entscheiden*”, which means “*I cannot decide*”. A word-by-word translation would be “*I can me not decide*”. The second example is “*wir bemühen uns, neue Arbeitsplätze zu schaffen*”. Its meaning is “*we try to create new jobs*”, and the word-by-word translation reads “*we try us, new jobs to create*”. In both cases, for each pair of consecutive words taken from the German source sentence all translation possibilities as found in the dictionary are given, and the translation selected by the system is highlighted.

[ich kann]	[can] [can i] [can * i] [cannot] [could] [i] [i am] [i am able] [i can] [i cannot] [i cannot * can] [i cannot * i * i] [i could] [i do] [i have] [i may] [i might] [i must] [i should] [i will] [i would] [i * able] [i * can] [i * cannot] [i * i] [i * i can] [i * is] [i * it] [in] [is] [is * i] [it] [it is] [it may] [let] [may] [me] [my] [my * is] [there is] [this is] [this will] [we] [we cannot] [what i can] [will]
[kann mich]	[am] [can] [can i] [can * i] [can * me] [can * myself] [cannot] [could] [do] [have] [i] [i can] [is] [may] [me] [will] [would]
[mich nicht]	[am] [am not] [am * not] [cannot] [do not] [do not * me] [does not] [does not * me] [have] [have not] [i] [i am] [i am not] [i cannot] [i do not] [i * not] [me] [me not] [me * not] [my] [my * not] [no] [no * me] [not] [not * am] [not * i] [not * i am] [not * me] [not * myself] [would not]
[nicht entscheiden]	[cannot decide] [decide] [not] [not * decide]

Table 5: Dictionary entries and translation results for the German sentence “*ich kann mich nicht entscheiden*”.

Note that the second example is more sophisticated, as the translated sentence cannot simply be constructed by concatenating the word pair translations. Instead, a reordering is necessary in this case. This leads to a further expansion of the search space, which forced us to introduce the limiting heuristic of giving priority to longer sequences. This means that we cannot guarantee that the entire space has been taken into account. We nevertheless hope that our examples give an impression how the system works, and what the problems are.

[wir bemühen]	[we] [we are] [we are endeavouring] [we are trying] [we seek] [we try] [we * trying]
[bemühen uns]	[are] [are endeavouring] [are trying] [are * trying] [have] [seek] [to] [try] [we]
[uns ,]	[,] [, we] [, * are] [, * have] [, * us] [, * we] [are] [are * ,] [are * that] [can] [do] [have] [have * ,] [our] [our * ,] [our * that] [ourselves] [ourselves ,] [that] [us] [us ,] [us -] [us and] [us that] [us which] [us * ,] [us * that] [us * we] [us * which] [we] [we ,] [we are] [we have] [we * ,] [what] [which] [will]
[, neue]	[,] [, new] [, * new] [and new] [fresh] [more] [new] [new * ,]
[neue arbeitsplätze]	[create employment] [create jobs] [create new jobs] [employment] [job creation] [jobs] [jobs * jobs] [more jobs] [new] [new employment] [new job] [new job opportunities] [new jobs] [new * employment] [new * jobs]
[arbeitsplätze zu]	[create jobs] [creation] [creation * jobs] [employment] [job] [jobs] [to] [to * employment] [to * jobs]
[zu schaffen]	['] [an] [are] [are * to] [at] [be] [can] [create] [create * to] [created] [creating] [creating * to] [creation] [establish] [for] [on] [on creating] [set] [should] [should be] [so] [to] [to achieve] [to be] [to bring] [to build] [to create] [to create * can] [to create * create] [to creating] [to do] [to ensure] [to establish] [to make] [to provide] [to * create] [to * created] [to * which] [which] [will] [will * to] [with] [would be]

Table 6: Dictionary entries and translation results for the German sentence “*wir bemühen uns, neue Arbeitsplätze zu schaffen*”.

5. Conclusions and Prospects

In a globalizing world the need for translations has been constantly increasing. If translation quality can be improved, machine translation can play an important role to satisfy this need. Self-learning systems that automatically derive their knowledge from sample translations are of particular interest as they make it possible to quickly adapt to new language pairs. In contrast, creating a new language pair for a rule-based system may take years of manual work.

The combinatorial approach to machine translation that we introduced here shows a way how highly informative bigram-based dictionaries can be used to address two core problems in machine translation, namely the problem of word ambiguity and the problem of word reordering.

Future work comprises the steps outlined in Table 2, with an emphasis on improving the translation engine. This includes looking for algorithms and heuristics to improve efficiency, but also to investigate if consideration of word sequences longer than pairs has a positive effect on translation quality.

6. Acknowledgements

We would like to thank Jörg Tiedemann for kindly providing his word alignments of the English/German part of the Europarl corpus, and the European Commission for financially supporting this work in the framework of the Marie Curie programme.

7. References

- Armstrong, S.; Kempen, M.; Petitpierre, D.; Rapp, R.; Thompson, H. (1998). Multilingual corpora for cooperation. *Proceedings of the 1st International Conference on Linguistic Resources and Evaluation (LREC), Granada*, Vol. 2, 975–980.
- Arnold, D.; Balkan, L.; Lee Humphreys, R.; Meijer, S.; Sadler, L. (1994). *Machine Translation. An Introductory Guide*. Manchester: NCC Blackwell.
- Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Lafferty, J., Mercer, R., Rossin, P. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2), 79–85.
- Fung, P.; Yee, L.Y. (1998). An IR approach for translating new words from nonparallel, comparable texts. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics, 1998*, Montreal, Vol. 1, 414–420.
- Gale, W.A.; Church, K.W. (1993). A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1), 75–102.
- Hutchins, J. (1987). Prospects in Machine Translation. In: M. Nagao (ed.): *Proceedings of the first MT Summit*, Tokyo, 7–12.
- Hutchins, W. John, Somers, Harold. L. (1992). *An Introduction to Machine Translation*. London: Academic Press.
- Koehn, P. (2002). *Europarl: A Multilingual Corpus for Evaluation of Machine Translation* (web publication). <http://people.csail.mit.edu/koehn/publications/europarl/>
- Koehn, P.; Knight, K. (2002). Learning a translation lexicon from monolingual corpora. In: *Proceedings of ACL-02 Workshop on Unsupervised Lexical Acquisition*, Philadelphia PA.
- Kay, M. (1995). Machine translation: The disappointing past and present. In: R.A. Cole; J. Mariani; H. Uszkoreit; A. Zaenen; V. Zue (eds.): *Survey of the State of the Art in Human Language Technology*, 285–300. URL: <http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>
- Och, F.J.; Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1), 19–51.
- Papineni, K.; Roukos, S.; Ward, T.; Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. In: *Proceedings of the 40th Annual Meeting of the ACL*, Philadelphia, PA, 311–318.
- Rabiner, L.R. (1990). A tutorial on Hidden Markov Models and selected applications in speech recognition. In: A. Waibel, K. Lee, (eds.): *Readings in Speech Recognition*. San Mateo: Morgan Kaufman. 267–296.
- Rapp, R. (1999). Automatic identification of word translations from unrelated English and German corpora. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics 1999, College Park, Maryland*. 519–526.
- Rapp, R. (2004a). A freely available automatically generated thesaurus of related words. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Vol. II, 395–398.
- Rapp, R. (2004b). A practical solution to the problem of automatic word sense induction. In: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, Companion Volume, 195–198.
- Sato, S.; Nagao, M. (1990). Toward memory-based translation. *Proceedings of COLING 1990*, 247–252.
- Slocum, J. (1989). Machine Translation: A Survey of Active Systems. In: In: I. Batori, W. Lenders, W. Putschke (eds.): *Computational Linguistics*. Berlin: de Gruyter, 629–645.
- Su, K.Y.; Chang, J.S. (1992). Why Corpus-Based Statistics-Oriented Machine Translation. In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Montreal, Canada, 1992, 249–262.
- Tiedemann, J. (2003). *Recycling Translations - Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing*. Doctoral Thesis, Studia Linguistica Upsaliensia 1.
- Weaver, W. (1949). Translation. Reprinted in: W. Locke, A. Booth (eds.) (1955): *Machine Translation of Languages*. Cambridge, Massachusetts: MIT Press, 15–23.