

# Automatic Identification of Word Translations from Unrelated English and German Corpora

Reinhard Rapp  
University of Mainz, FASK  
D-76711 Gernersheim, Germany  
rapp@usun2.fask.uni-mainz.de

## Abstract

Algorithms for the alignment of words in translated texts are well established. However, only recently new approaches have been proposed to identify word translations from non-parallel or even unrelated texts. This task is more difficult, because most statistical clues useful in the processing of parallel texts cannot be applied to non-parallel texts. Whereas for parallel texts in some studies up to 99% of the word alignments have been shown to be correct, the accuracy for non-parallel texts has been around 30% up to now. The current study, which is based on the assumption that there is a correlation between the patterns of word co-occurrences in corpora of different languages, makes a significant improvement to about 72% of word translations identified correctly.

## 1 Introduction

Starting with the well-known paper of Brown et al. (1990) on statistical machine translation, there has been much scientific interest in the alignment of sentences and words in translated texts. Many studies show that for nicely parallel corpora high accuracy rates of up to 99% can be achieved for both sentence and word alignment (Gale & Church, 1993; Kay & Röscheisen, 1993). Of course, in practice – due to omissions, transpositions, insertions, and replacements in the process of translation – with real texts there may be all kinds of problems, and therefore robustness is still an issue (Langlais et al., 1998).

Nevertheless, the results achieved with these algorithms have been found useful for the com-

pilation of dictionaries, for checking the consistency of terminological usage in translations, for assisting the terminological work of translators and interpreters, and for example-based machine translation. By now, some alignment programs are offered commercially: Translation memory tools for translators, such as IBM's Translation Manager or Trados' Translator's Workbench, are bundled or can be upgraded with programs for sentence alignment.

Most of the proposed algorithms first conduct an alignment of sentences, that is, they locate those pairs of sentences that are translations of each other. In a second step a word alignment is performed by analyzing the correspondences of words in each pair of sentences. The algorithms are usually based on one or several of the following statistical clues:

1. correspondence of word and sentence order
2. correlation between word frequencies
3. cognates: similar spelling of words in related languages

All these clues usually work well for parallel texts. However, despite serious efforts in the compilation of parallel corpora (Armstrong et al., 1998), the availability of a large-enough parallel corpus in a specific domain and for a given pair of languages is still an exception. Since the acquisition of monolingual corpora is much easier, it would be desirable to have a program that can determine the translations of words from comparable (same domain) or possibly unrelated monolingual texts of two languages. This is what translators and interpreters usually do when preparing terminology in a specific field: They read texts corresponding to this field in both languages and draw their conclusions on word correspondences from the usage of the

terms. Of course, the translators and interpreters can understand the texts, whereas our programs are only considering a few statistical clues.

For non-parallel texts the first clue, which is usually by far the strongest of the three mentioned above, is not applicable at all. The second clue is generally less powerful than the first, since most words are ambiguous in natural languages, and many ambiguities are different across languages. Nevertheless, this clue is applicable in the case of comparable texts, although with a lower reliability than for parallel texts. However, in the case of unrelated texts, its usefulness may be near zero. The third clue is generally limited to the identification of word pairs with similar spelling. For all other pairs, it is usually used in combination with the first clue. Since the first clue does not work with non-parallel texts, the third clue is useless for the identification of the majority of pairs. For unrelated languages, it is not applicable anyway.

In this situation, Rapp (1995) proposed using a clue different from the three mentioned above: His co-occurrence clue is based on the assumption that there is a correlation between co-occurrence patterns in different languages. For example, if the words *teacher* and *school* co-occur more often than expected by chance in a corpus of English, then the German translations of *teacher* and *school*, *Lehrer* and *Schule*, should also co-occur more often than expected in a corpus of German. In a feasibility study he showed that this assumption actually holds for the language pair English/German even in the case of unrelated texts. When comparing an English and a German co-occurrence matrix of corresponding words, he found a high correlation between the co-occurrence patterns of the two matrices when the rows and columns of both matrices were in corresponding word order, and a low correlation when the rows and columns were in random order.

The validity of the co-occurrence clue is obvious for parallel corpora, but – as empirically shown by Rapp – it also holds for non-parallel corpora. It can be expected that this clue will work best with parallel corpora, second-best with comparable corpora, and somewhat worse with unrelated corpora. In all three cases, the problem of robustness – as observed when

applying the word-order clue to parallel corpora – is not severe. Transpositions of text segments have virtually no negative effect, and omissions or insertions are not critical. However, the co-occurrence clue when applied to comparable corpora is much weaker than the word-order clue when applied to parallel corpora, so larger corpora and well-chosen statistical methods are required.

After an attempt with a context heterogeneity measure (Fung, 1995) for identifying word translations, Fung based her later work also on the co-occurrence assumption (Fung & Yee, 1998; Fung & McKeown, 1997). By presupposing a lexicon of seed words, she avoids the prohibitively expensive computational effort encountered by Rapp (1995). The method described here – although developed independently of Fung's work – goes in the same direction. Conceptually, it is a trivial case of Rapp's matrix permutation method. By simply assuming an initial lexicon the large number of permutations to be considered is reduced to a much smaller number of vector comparisons. The main contribution of this paper is to describe a practical implementation based on the co-occurrence clue that yields good results.

## 2 Approach

As mentioned above, it is assumed that across languages there is a correlation between the co-occurrences of words that are translations of each other. If – for example – in a text of one language two words *A* and *B* co-occur more often than expected by chance, then in a text of another language those words that are translations of *A* and *B* should also co-occur more frequently than expected. This is the only statistical clue used throughout this paper.

It is further assumed that there is a small dictionary available at the beginning, and that our aim is to expand this base lexicon. Using a corpus of the target language, we first compute a co-occurrence matrix whose rows are all word types occurring in the corpus and whose columns are all target words appearing in the base lexicon. We now select a word of the source language whose translation is to be determined. Using our source-language corpus, we compute

a co-occurrence vector for this word. We translate all known words in this vector to the target language. Since our base lexicon is small, only some of the translations are known. All unknown words are discarded from the vector and the vector positions are sorted in order to match the vectors of the target-language matrix. With the resulting vector, we now perform a similarity computation to all vectors in the co-occurrence matrix of the target language. The vector with the highest similarity is considered to be the translation of our source-language word.

### 3 Simulation

#### 3.1 Language Resources

To conduct the simulation, a number of resources were required. These are

1. a German corpus
2. an English corpus
3. a number of German test words with known English translations
4. a small base lexicon, German to English

As the German corpus, we used 135 million words of the newspaper *Frankfurter Allgemeine Zeitung* (1993 to 1996), and as the English corpus 163 million words of the *Guardian* (1990 to 1994). Since the orientation of the two newspapers is quite different, and since the time spans covered are only in part overlapping, the two corpora can be considered as more or less unrelated.

For testing our results, we started with a list of 100 German test words as proposed by Russell (1970), which he used for an association experiment with German subjects. By looking up the translations for each of these 100 words, we obtained a test set for evaluation.

Our German/English base lexicon is derived from the *Collins Gem German Dictionary* with about 22,300 entries. From this we eliminated all multi-word entries, so 16,380 entries remained. Because we had decided on our test word list beforehand, and since it would not make much sense to apply our method to words that are already in the base lexicon, we also removed all entries belonging to the 100 test words.

#### 3.2 Pre-processing

Since our corpora are very large, to save disk space and processing time we decided to remove all function words from the texts. This was done on the basis of a list of approximately 600 German and another list of about 200 English function words. These lists were compiled by looking at the closed class words (mainly articles, pronouns, and particles) in an English and a German morphological lexicon (for details see Lezius, Rapp, & Wettler, 1998) and at word frequency lists derived from our corpora.<sup>1</sup> By eliminating function words, we assumed we would lose little information: Function words are often highly ambiguous and their co-occurrences are mostly based on syntactic instead of semantic patterns. Since semantic patterns are more reliable than syntactic patterns across language families, we hoped that eliminating the function words would give our method more generality.

We also decided to lemmatize our corpora. Since we were interested in the translations of base forms only, it was clear that lemmatization would be useful. It not only reduces the sparse-data problem but also takes into account that German is a highly inflectional language, whereas English is not. For both languages we conducted a partial lemmatization procedure that was based only on a morphological lexicon and did not take the context of a word form into account. This means that we could not lemmatize those ambiguous word forms that can be derived from more than one base form. However, this is a relatively rare case. (According to Lezius, Rapp, & Wettler, 1998, 93% of the tokens of a German text had only one lemma.) Although we had a context-sensitive lemmatizer for German available (Lezius, Rapp, & Wettler, 1998), this was not the case for English, so for reasons of symmetry we decided not to use the context feature.

---

<sup>1</sup> In cases in which an ambiguous word can be both a content and a function word (e.g., *can*), preference was given to those interpretations that appeared to occur more frequently.

### 3.3 Co-occurrence Counting

For counting word co-occurrences, in most other studies a fixed window size is chosen and it is determined how often each pair of words occurs within a text window of this size. However, this approach does not take word order within a window into account. Since it has been empirically observed that word order of content words is often similar between languages (even between unrelated languages such as English and Chinese), and since this may be a useful statistical clue, we decided to modify the common approach in the way proposed by Rapp (1996, p. 162). Instead of computing a single co-occurrence vector for a word *A*, we compute several, one for each position within the window. For example, if we have chosen the window size 2, we would compute a first co-occurrence vector for the case that word *A* is two words ahead of another word *B*, a second vector for the case that word *A* is one word ahead of word *B*, a third vector for *A* directly following *B*, and a fourth vector for *A* following two words after *B*. If we added up these four vectors, the result would be the co-occurrence vector as obtained when not taking word order into account. However, this is not what we do. Instead, we combine the four vectors of length *n* into a single vector of length *4n*.

Since preliminary experiments showed that a window size of 3 with consideration of word order seemed to give somewhat better results than other window types, the results reported here are based on vectors of this kind. However, the computational methods described below are in the same way applicable to window sizes of any length with or without consideration of word order.

### 3.4 Association Formula

Our method is based on the assumption that there is a correlation between the patterns of word co-occurrences in texts of different languages. However, as Rapp (1995) proposed, this correlation may be strengthened by not using the co-occurrence counts directly, but association strengths between words instead. The idea is to eliminate word-frequency effects and to emphasize significant word pairs by comparing their

observed co-occurrence counts with their expected co-occurrence counts. In the past, for this purpose a number of measures have been proposed. They were based on mutual information (Church & Hanks, 1989), conditional probabilities (Rapp, 1996), or on some standard statistical tests, such as the chi-square test or the log-likelihood ratio (Dunning, 1993). For the purpose of this paper, we decided to use the log-likelihood ratio, which is theoretically well justified and more appropriate for sparse data than chi-square. In preliminary experiments it also led to slightly better results than the conditional probability measure. Results based on mutual information or co-occurrence counts were significantly worse. For efficient computation of the log-likelihood ratio we used the following formula:<sup>2</sup>

$$\begin{aligned} -2 \log \lambda &= \sum_{i,j \in \{1,2\}} k_{ij} \log \frac{k_{ij}N}{C_i R_j} \\ &= k_{11} \log \frac{k_{11}N}{C_1 R_1} + k_{12} \log \frac{k_{12}N}{C_1 R_2} \\ &\quad + k_{21} \log \frac{k_{21}N}{C_2 R_1} + k_{22} \log \frac{k_{22}N}{C_2 R_2} \end{aligned}$$

where

$$\begin{aligned} C_1 &= k_{11} + k_{12} & C_2 &= k_{21} + k_{22} \\ R_1 &= k_{11} + k_{21} & R_2 &= k_{12} + k_{22} \\ N &= k_{11} + k_{12} + k_{21} + k_{22} \end{aligned}$$

with parameters  $k_{ij}$  expressed in terms of corpus frequencies:

- $k_{11}$  = frequency of common occurrence of word *A* and word *B*
- $k_{12}$  = corpus frequency of word *A* –  $k_{11}$
- $k_{21}$  = corpus frequency of word *B* –  $k_{11}$
- $k_{22}$  = size of corpus (no. of tokens) – corpus frequency of *A* – corpus frequency of *B*

All co-occurrence vectors were transformed using this formula. Thereafter, they were normalized in such a way that for each vector the sum of its entries adds up to one. In the rest of the paper, we refer to the transformed and normalized vectors as association vectors.

<sup>2</sup> This formulation of the log-likelihood ratio was proposed by Ted Dunning during a discussion on the corpora mailing list (e-mail of July 22, 1997). It is faster and more mnemonic than the one in Dunning (1993).

### 3.5 Vector Similarity

To determine the English translation of an unknown German word, the association vector of the German word is computed and compared to all association vectors in the English association matrix. For comparison, the correspondences between the vector positions and the columns of the matrix are determined by using the base lexicon. Thus, for each vector in the English matrix a similarity value is computed and the English words are ranked according to these values. It is expected that the correct translation is ranked first in the sorted list.

For vector comparison, different similarity measures can be considered. Salton & McGill (1983) proposed a number of measures, such as the Cosine coefficient, the Jaccard coefficient, and the Dice coefficient (see also Jones & Furnas, 1987). For the computation of related terms and synonyms, Ruge (1995), Landauer and Dumais (1997), and Fung and McKeown (1997) used the cosine measure, whereas Grefenstette (1994, p. 48) used a weighted Jaccard measure. We propose here the city-block metric, which computes the similarity between two vectors  $X$  and  $Y$  as the sum of the absolute differences of corresponding vector positions:

$$s = \sum_{i=1}^n |X_i - Y_i|$$

In a number of experiments we compared it to other similarity measures, such as the cosine measure, the Jaccard measure (standard and binary), the Euclidean distance, and the scalar product, and found that the city-block metric yielded the best results. This may seem surprising, since the formula is very simple and the computational effort smaller than with the other measures. It must be noted, however, that the other authors applied their similarity measures directly to the (log of the) co-occurrence vectors, whereas we applied the measures to the association vectors based on the log-likelihood ratio. According to our observations, estimates based on the log-likelihood ratio are generally more reliable across different corpora and languages.

### 3.6 Simulation Procedure

The results reported in the next section were obtained using the following procedure:

1. Based on the word co-occurrences in the German corpus, for each of the 100 German test words its association vector was computed. In these vectors, all entries belonging to words not found in the English part of the base lexicon were deleted.
2. Based on the word co-occurrences in the English corpus, an association matrix was computed whose rows were all word types of the corpus with a frequency of 100 or higher<sup>3</sup> and whose columns were all English words occurring as first translations of the German words in the base lexicon.<sup>4</sup>
3. Using the similarity function, each of the German vectors was compared to all vectors of the English matrix. The mapping between vector positions was based on the first translations given in the base lexicon. For each of the German source words, the English vocabulary was ranked according to the resulting similarity value.

---

<sup>3</sup> The limitation to words with frequencies above 99 was introduced for computational reasons to reduce the number of vector comparisons and thus speed up the program. (The English corpus contains 657,787 word types after lemmatization, which leads to extremely large matrices.) The purpose of this limitation was not to limit the number of translation candidates considered. Experiments with lower thresholds showed that this choice has little effect on the results to our set of test words.

<sup>4</sup> This means that alternative translations of a word were not considered. Another approach, as conducted by Fung & Yee (1998), would be to consider all possible translations listed in the lexicon and to give them equal (or possibly descending) weight. Our decision was motivated by the observation that many words have a salient first translation and that this translation is listed first in the Collins Gem Dictionary German-English. We did not explore this issue further since in a small pocket dictionary only few ambiguities are listed.

## 4 Results and Evaluation

Table 1 shows the results for 20 of the 100 German test words. For each of these test words, the top five translations as automatically generated are listed. In addition, for each word its expected English translation from the test set is given together with its position in the ranked lists of computed translations. The positions in the ranked lists are a measure for the quality of the predictions, with a 1 meaning that the prediction is correct and a high value meaning that the program was far from predicting the correct word.

If we look at the table, we see that in many cases the program predicts the expected word, with other possible translations immediately following. For example, for the German word *Häuschen*, the correct translations *bungalow*, *cottage*, *house*, and *hut* are listed. In other cases, typical associates follow the correct translation. For example, the correct translation of *Mädchen*, *girl*, is followed by *boy*, *man*, *brother*, and *lady*. This behavior can be expected from our associationist approach. Unfortunately, in some cases the correct translation and one of its strong associates are mixed up, as for example with *Frau*, where its correct translation, *woman*, is listed only second after its strong associate *man*. Another example of this typical kind of error is *pfeifen*, where the correct translation *whistle* is listed third after *linesman* and *referee*.

Let us now look at some cases where the program did particularly badly. For *Kohl* we had expected its dictionary translation *cabbage*, but – given that a substantial part of our newspaper corpora consists of political texts – we do not need to further explain why our program lists *Major*, *Kohl*, *Thatcher*, *Gorbachev*, and *Bush*, state leaders who were in office during the time period the texts were written. In other cases, such as *Krankheit* and *Whisky*, the simulation program simply preferred the British usage of the Guardian over the American usage in our test set: Instead of *sickness*, the program predicted *disease* and *illness*, and instead of *whiskey* it predicted *whisky*.

A much more severe problem is that our current approach cannot properly handle ambiguities: For the German word *weiß* it does not pre-

dict *white*, but instead *know*. The reason is that *weiß* can also be third person singular of the German verb *wissen* (to know), which in newspaper texts is more frequent than the color *white*. Since our lemmatizer is not context-sensitive, this word was left unlemmatized, which explains the result.

To be able to compare our results with other work, we also did a quantitative evaluation. For all test words we checked whether the predicted translation (first word in the ranked list) was identical to our expected translation. This was true for 65 of the 100 test words. However, in some cases the choice of the expected translation in the test set had been somewhat arbitrary. For example, for the German word *Straße* we had expected *street*, but the system predicted *road*, which is a translation quite as good. Therefore, as a better measure for the accuracy of our system we counted the number of times where an acceptable translation of the source word is ranked first. This was true for 72 of the 100 test words, which gives us an accuracy of 72%. In another test, we checked whether an acceptable translation appeared among the top 10 of the ranked lists. This was true in 89 cases.<sup>5</sup>

For comparison, Fung & McKeown (1997) report an accuracy of about 30% when only the top candidate is counted. However, it must be emphasized that their result has been achieved under very different circumstances. On the one hand, their task was more difficult because they worked on a pair of unrelated languages (English/Japanese) using smaller corpora and a random selection of test words, many of which were multi-word terms. Also, they predetermined a single translation as being correct. On the other hand, when conducting their evaluation, Fung & McKeown limited the vocabulary they considered as translation candidates to a few hundred terms, which obviously facilitates the task.

---

<sup>5</sup> We did not check for the completeness of the translations found (recall), since this measure depends very much on the size of the dictionary used as the standard.

German test word	expected translation and rank	top five translations as automatically generated					
Baby	baby 1	baby	child	mother	daughter	father	
Brot	bread 1	bread	cheese	meat	food	butter	
Frau	woman 2	man	woman	boy	friend	wife	
gelb	yellow 1	yellow	blue	red	pink	green	
Häuschen	cottage 2	bungalow	cottage	house	hut	village	
Kind	child 1	child	daughter	son	father	mother	
Kohl	cabbage 17074	Major	Kohl	Thatcher	Gorbachev	Bush	
Krankheit	sickness 86	disease	illness	Aids	patient	doctor	
Mädchen	girl 1	girl	boy	man	brother	lady	
Musik	music 1	music	theatre	musical	dance	song	
Ofen	stove 3	heat	oven	stove	house	burn	
pfeifen	whistle 3	linesman	referee	whistle	blow	offside	
Religion	religion 1	religion	culture	faith	religious	belief	
Schaf	sheep 1	sheep	cattle	cow	pig	goat	
Soldat	soldier 1	soldier	army	troop	force	civilian	
Straße	street 2	road	street	city	town	walk	
süß	sweet 1	sweet	smell	delicious	taste	love	
Tabak	tobacco 1	tobacco	cigarette	consumption	nicotine	drink	
weiß	white 46	know	say	thought	see	think	
Whisky	whiskey 11	whisky	beer	Scotch	bottle	wine	

Table 1: Results for 20 of the 100 test words (for full list see <http://www.fask.uni-mainz.de/user/rapp/>)

## 5 Discussion and Conclusion

The method described can be seen as a simple case of the gradient descent method proposed by Rapp (1995), which does not need an initial lexicon but is computationally prohibitively expensive. It can also be considered as an extension from the monolingual to the bilingual case of the well-established methods for semantic or syntactic word clustering as proposed by Schütze (1993), Grefenstette (1994), Ruge (1995), Rapp (1996), Lin (1998), and others. Some of these authors perform a shallow or full syntactical analysis before constructing the co-occurrence vectors. Others reduce the size of the co-occurrence matrices by performing a singular value decomposition. However, in yet unpublished work we found that at least for the computation of synonyms and related words neither syntactical analysis nor singular value decomposition lead to significantly better results than the approach described here when applied to the monolingual case (see also Grefenstette,

1993), so we did not try to include these methods in our system. Nevertheless, both methods are of technical value since they lead to a reduction in the size of the co-occurrence matrices.

Future work has to approach the difficult problem of ambiguity resolution, which has not been dealt with here. One possibility would be to semantically disambiguate the words in the corpora beforehand, another to look at co-occurrences between significant word sequences instead of co-occurrences between single words.

To conclude with, let us add some speculation by mentioning that the ability to identify word translations from non-parallel texts can be seen as an indicator in favor of the associationist view of human language acquisition (see also Landauer & Dumais, 1997, and Wettler & Rapp, 1993). It gives us an idea of how it is possible to derive the meaning of unknown words from texts by only presupposing a limited number of known words and then iteratively expanding this knowledge base. One possibility to get the pro-

cess going would be to learn vocabulary lists as in school, another to simply acquire the names of items in the physical world.

### Acknowledgements

I thank Manfred Wettler, Gisela Zunker-Rapp, Wolfgang Lezius, and Anita Todd for their support of this work.

### 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.
- Brown, P.; Cocke, J.; Della Pietra, S. A.; Della Pietra, V. J.; Jelinek, F.; Lafferty, J. D.; Mercer, R. L.; Rossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2), 79-85.
- Church, K. W.; Hanks, P. (1989). Word association norms, mutual information, and lexicography. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver, British Columbia, 76-83.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1), 61-74.
- Fung, P. (1995). Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus. *Proceedings of the 3rd Annual Workshop on Very Large Corpora*, Boston, Massachusetts, 173-183.
- Fung, P.; McKeown, K. (1997). Finding terminology translations from non-parallel corpora. *Proceedings of the 5th Annual Workshop on Very Large Corpora*, Hong Kong, 192-202.
- Fung, P.; Yee, L. Y. (1998). An IR approach for translating new words from nonparallel, comparable texts. In: *Proceedings of COLING-ACL 1998*, Montreal, Vol. 1, 414-420.
- Gale, W. A.; Church, K. W. (1993). A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(3), 75-102.
- Grefenstette, G. (1993). Evaluation techniques for automatic semantic extraction: comparing syntactic and window based approaches. In: *Proceedings of the Workshop on Acquisition of Lexical Knowledge from Text*, Columbus, Ohio.
- Grefenstette, G. (1994). *Explorations in Automatic Thesaurus Discovery*. Dordrecht: Kluwer.
- Jones, W. P.; Furnas, G. W. (1987). Pictures of relevance: a geometric analysis of similarity measures. *Journal of the American Society for Information Science*, 38(6), 420-442.
- Kay, M.; Röscheisen, M. (1993). Text-Translation Alignment. *Computational Linguistics*, 19(1), 121-142.
- Landauer, T. K.; Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211-240.
- Langlais, P.; Simard, M.; Véronis, J. (1998). Methods and practical issues in evaluating alignment techniques. In: *Proceedings of COLING-ACL 1998*, Montreal, Vol. 1, 711-717.
- Lezius, W.; Rapp, R.; Wettler, M. (1998). A freely available morphology system, part-of-speech tagger, and context-sensitive lemmatizer for German. In: *Proceedings of COLING-ACL 1998*, Montreal, Vol. 2, 743-748.
- Lin, D. (1998). Automatic Retrieval and Clustering of Similar Words. In: *Proceedings of COLING-ACL 1998*, Montreal, Vol. 2, 768-773.
- Rapp, R. (1995). Identifying word translations in non-parallel texts. In: *Proceedings of the 33rd Meeting of the Association for Computational Linguistics*. Cambridge, Massachusetts, 320-322.
- Rapp, R. (1996). *Die Berechnung von Assoziationen*. Hildesheim: Olms.
- Ruge, G. (1995). Human memory models and term association. *Proceedings of the ACM SIGIR Conference*, Seattle, 219-227.
- Russell, W. A. (1970). The complete German language norms for responses to 100 words from the Kent-Rosanoff word association test. In: L. Postman, G. Keppel (eds.): *Norms of Word Association*. New York: Academic Press, 53-94.
- Salton, G.; McGill, M. (1983). *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.
- Schütze, H. (1993). Part-of-speech induction from scratch. In: *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, 251-258.
- Wettler, M.; Rapp, R. (1993). Computation of word associations based on the co-occurrences of words in large corpora. In: *Proceedings of the 1st Workshop on Very Large Corpora*: Columbus, Ohio, 84-93.