

Combining Clues for Word Alignment

Jörg Tiedemann

Department of Linguistics

Uppsala University

Box 527

SE-751 20 Uppsala, Sweden

joerg@stp.ling.uu.se

Abstract

In this paper, a word alignment approach is presented which is based on a combination of clues. Word alignment clues indicate associations between words and phrases. They can be based on features such as frequency, part-of-speech, phrase type, and the actual wordform strings. Clues can be found by calculating similarity measures or learned from word aligned data. The *clue alignment* approach, which is proposed in this paper, makes it possible to combine association clues taking different kinds of linguistic information into account. It allows a dynamic tokenization into token units of varying size. The approach has been applied to an English/Swedish parallel text with promising results.

1 Introduction

Parallel corpora carry a huge amount of bilingual lexical information. Word alignment approaches focus on the automatic identification of translation relations in translated texts. Alignments are usually represented as a set of links between words and phrases of source and target language segments. An alignment can be complete, i.e. all items in both segments have been linked to corresponding items in the other language, or incomplete, otherwise. Alignments may include “null

links” which can be modeled as links to an “empty element”.

In word alignment, we have to

- find an appropriate model M for the alignment of source and target language texts (**modeling**)
- estimate parameters of the model M , e.g. from empirical data (**parameter estimation**)
- find the optimal alignment of words and phrases for a given translation according to the model M and its parameters (**alignment recovery**).

Modeling the relations between lexical units of translated texts is not a trivial task due to the diversity of natural languages. There are generally two approaches, the *estimation approach* which is used in, e.g., statistical machine translation, and the *association approach* which is used in, e.g., automatic extraction of bilingual terminology. In the estimation approach, alignment parameters are modeled as hidden parameters in a statistical translation model (Och and Ney, 2000). Association approaches base the alignment on similarity measures and association tests such as Dice scores (Smadja et al., 1996; Tiedemann, 1999), t-scores (Ahrenberg et al., 1998) log-likelihood measures (Tufis and Barbu, 2002), and longest common subsequence ratios (Melamed, 1995).

One of the main difficulties in all alignment strategies is the identification of appropriate units in the source and the target language to be aligned. This task is hard even for human experts as can be seen in the detailed guidelines which are required

for manual alignments (Merkel, 1999; Melamed, 1998). Many translation relations involve multi-word units such as phrasal compounds, idiomatic expressions, and complex terms. Syntactic shifts can also require the consideration of a context larger than a single word. Some items are not translated at all. Splitting source and target language texts into appropriate units for alignment (henceforth: *tokenization*) is often not possible without considering the translation relations. In other words, initial tokenization borders may change when the translation relations are investigated. Human aligners frequently expand token units when aligning sentences manually depending on the context (Ahrenberg et al., 2002). Previous approaches use either iterative procedures to re-estimate alignment parameters (Smadja et al., 1996; Melamed, 1997; Vogel et al., 2000) or pre-processing steps for the identification of token N-grams (Ahrenberg et al., 1998; Tiedemann, 1999). In our approach, we combine simple techniques for prior tokenization with dynamic techniques during the alignment phase.

The second problem of traditional word alignment approaches is the fact that parameter estimations are usually based on plain text items only. Linguistic data, which could be used to identify associations between lexical items are often ignored. Linguistic tools such as part-of-speech taggers, (shallow) parsers, named-entity recognizers become more and more robust and available for more languages. Linguistic information including contextual features could be used to improve alignment strategies.

The third problem, alignment recovery, is a search problem. Using the alignment model and its parameters, we have to find the optimal alignment for a given pair of source and target language segments. In (Hiemstra, 1998), the author points out that a sentence pair with a maximum of n token units in both sentences has $n!$ possible alignments in a simple directed alignment model with a fixed tokenization. Furthermore, a search strategy becomes very complex if we allow dynamic tokenization borders (overlapping N-grams, inclusions), which leads us not only to a larger number of possible combinations but also to the problem of comparing alignments with variable

length (number of links).

The *clue alignment approach*, which we propose here, addresses the three problems which were mentioned above. The approach allows the combination of association measures for any features of translation units of varying size. Overlapping units are allowed as well as inclusions. Association scores are organized in a *clue matrix* and we present a simple approach for approximating the optimal alignment.

Section 2 describes the clue alignment model and ways of estimating parameters from association scores. Section 3 introduces the alignment approach which is based on word alignment clues. Section 4 gives examples of learning clues from previous alignments. Section 5 summarizes alignment experiments and, finally, section 6 contains conclusions and a discussion.

2 Word Alignment Clues

2.1 Motivation

The following English/Swedish sentence pair has been taken from the PLUG corpus (Sågvald Hein, 1999):

The corridors are jumping with them.
Korridorerna myllrar av dem.

The task for an aligner is now to find all the links between the lexical items in English and the lexical items in Swedish. The natural way of doing this for a human is to use various kinds of information, clues. Even without knowing either of the two languages, a human aligner would find a strong similarity between *corridors* and *korridorerna* which leads to the conclusion of a possible relation between these two words. Similarly, a relation could be seen between *them* and *dem*.¹

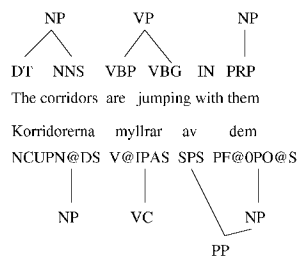
In a second step, the aligner might use frequency counts of words in both languages and co-occurrence frequencies for some interesting word pairs.

source	freq	target	freq	co-occ.	Dice
corridors	3	korridorerna	2	2	0.8
with	518	dem	188	39	0.110
with	518	av	988	169	0.224
them	193	dem	188	170	0.892
them	193	av	988	71	0.120

¹Of course, the aligner should be aware of the possibilities of false friends especially among short words.

The frequency table above gives the aligner an additional clue for an association between *corridors* and *korridorerna* and also some ideas about the relation between *them* and *dem* but not much about the remaining words.

Finally, the aligner might apply off-the-shelf part-of-speech taggers and shallow syntactic analyzers.



The aligner might look up the descriptions of the English tag set and finds out that *NNS* is the label for a plural noun, *VBP* and *VBG* are labels of verbs in the present tense, *IN* labels a preposition, and *PRP* a personal pronoun. Similarly, (s)he looks for the Swedish tags and finds out that *NCUPN@DS* describes a definite noun in plural form and nominative case, *V@IPAS* describes an active verb in the present tense, *SPS* labels a preposition, and *PF@OPO@S* describes a definite pronoun object in plural form. This gives the aligner additional clues about possible links. (S)he might expect relations between active verbs in the present tense rather than between verbs and nouns. Finding out that Swedish nouns can bear the feature of definiteness gives the aligner another clue about the translation of the definite article in the English sentence.

Finally, the aligner looks at the output of the shallow parser and gets additional clues for aligning the two sentences. For example, the two English verbs build a verb phrase (*VP*) which is most likely to be linked to the only “verb cluster” (*VC*) in the Swedish sentence. The personal pronoun in the English sentence is used as a noun phrase (*NP*) similar to the pronoun in the Swedish sentence.

Putting all the clues together, the aligner comes up with the following alignment without actually having to know the two languages:

the corridors	korridorerna
are jumping	myllrar
with	av
them	dem

However, looking at the sentence pair again, a second aligner with knowledge of both languages might realize that the verbs *myllrar* (English: swarm) and *jumping* do not really correspond to each other in isolation and that the expressions are rather idiomatic in both languages. Therefore, the second aligner might decide to link the whole expression “*are jumping with*” to the Swedish translation of “*myllrar av*”.

This kind of disagreement between human aligners is quite normal and demonstrates quite well the problems which have to be handled by automatic alignment approaches.

2.2 Definitions

Now, we would like to use a similar strategy as described in the previous section for an automatic alignment process. In our approach, we use the following definitions:

Word alignment clue: A word alignment clue $C_i(s, t)$ is a probability which indicates an association between two lexical items s and t in parallel texts.

Lexical item: A lexical item is a set of words with associated features attached to it (word position may be a feature).

A clue is called **static** if its value is constant for a given pair of features of lexical items, otherwise it is called **dynamic**. Furthermore, clues can be **declarative**, i.e. pre-defined feature correspondences, or **estimated**, i.e. from association scores or from training data. Generally, a clue is defined as a weighted association A between s and t :

$$C_i(s, t) = P(a_i) = w_i A_i(s, t)$$

The value of w_i is used to normalize and weight the association score A_i .

Alignment clues can be estimated from association measures given empirical data. Examples of such measures are given below:

Co-occurrence: $A_{Dice}(s, t) = \frac{2P(s,t)}{P(s)+P(t)}$
(the Dice coefficient)

String similarity: $A_{LCSR}(s, t) = LCSR(s, t)$
(the longest common subsequence ratio)

Other clues can be estimated from word aligned training data:

$$C_j(s, t) = w_j * P(f_t | f_s) \approx w_j \frac{\text{freq}(f_s, f_t)}{\text{freq}(f_s)}$$

f_s and f_t are sets of features of s and t , respectively. They may include features such as part-of-speech, phrase categories, word positions, and/or any other kind of contextual features.

Clues can also be pre-defined. For example, machine-readable dictionary can be used as a collection of declarative clues. Each translation from the dictionary is an alignment clue for the corresponding word pairs. The likelihood of each translation alternative can be weighted, e.g., by frequency (if available).

2.3 Clue Combinations

So far, word alignment clues are simply sets of weighted association scores. The key task is to combine available clues in order to find interlingual links. Clues are defined as probabilities of associations. In order to combine all indications which are given by single clues $C_i(s, t) = P(a_i)$ we define the overall clue $C_{all}(s, t)$ for a given pair of lexical items as the disjunction of all indications:

$$C_{all}(s, t) = P(a_{all}) = P(a_1 \cup a_2 \cup \dots \cup a_n)$$

Note that clues are not mutually exclusive. For example, an association based on co-occurrence measures can be found together with an association based on string similarity measures. Using the addition rule for probabilities we get the following formula for a disjunction of two clues:

$$P(a_1 \cup a_2) = P(a_1) + P(a_2) - P(a_1 \cap a_2)$$

For simplicity, we assume that clues are independent of each other.

$$P(a_1 \cap a_2) = P(a_1)P(a_2)$$

This is a crucial assumption and has to be considered when designing clue patterns.

2.4 Overlaps, Inclusions and the Clue Matrix

Word alignment clues may refer to any set of words from the source and target language segment according to the definitions in section 2.2. Therefore, clues can refer to sets of words which overlap with other sets of words to which another clue refers. Such overlaps and inclusions make it impossible to combine the corresponding clues directly with the formulas which were given in the previous section. In order to enable clue combi-

nations even for overlapping units, we define the following property of word alignment clues:

A clue indicates an association between all its member token pairs.

This property makes it possible to combine alignment clues by distributing the clue indication from complex structures to single word pairs. In this way, dynamic tokenization can be used for both, source and target language sentences and combined association scores (the total clue value) can be calculated for each pair of single tokens.

Now, sentence pairs can be represented in a two-dimensional matrix with one source language word per row and one target language word per column. The cells inside the matrix can be filled with the combined clue values for the corresponding word pairs. Henceforth, this matrix will be referred to as a *clue matrix*.

2.5 Example

Consider the following English/Swedish sentence pair:

Then hand baggage is opened.
Sedan öppnas handbagaget.

Assume that the alignment program found the following alignment clues which are based on string similarity and co-occurrence statistics:²

co-occurrence (DICE)		
then	sedan	0.38
is opened	öppnas	0.65
is opened	sedan öppnas	0.2
opened	öppnas	0.5
baggage	handbagaget	0.45
string similarity (LCSR)		
hand baggage	handbagaget	0.83
opened	öppnas	0.33
then	sedan	0.4
hand	sedan	0.4

The alignment clues contain only three multi-word units. However, even these few units cause several overlaps. For example, the English string “*hand baggage*” from the set of string similarity clues overlaps with the string “*baggage*”. The clue for the pair “*is opened*” and “*sedan öppnas*” overlaps with six other clues. However, using our

²Note that clues do not have to be correct! Alignment clues give hints for a possible relation between words and phrases. They can even be misleading, but hopefully, the indication of combined clues will lead to correct links.

definitions of alignment clues, we can easily construct the following clue matrix:

	sedan	öppnas	handbagaget
then	0.628	0	0
hand	0.4	0	0.83
baggage	0	0	0.9065
is	0.2	0.72	0
opened	0.2	0.86	0

The matrix is simply filled with all values of combined clues for each word pair. For example, the total clue value for the word pair $s = \text{“baggage”}$ and $t = \text{“handbagaget”}$ is calculated as follows:

$$C_{all}(s, t) = 0.45 + 0.83 - 0.45 * 0.83 = 0.9065$$

All other values are computed in the same way. Looking at the matrix, we can find clear relations between certain words such as *[hand, baggage]* and *handbagaget*. However, between other word pairs such as *is* and *sedan* we find only low associations which conflict with others and therefore, they can be dismissed in the alignment process.

3 Clue Alignment

Word alignment clues as described above can be used to model the relations between words of translated texts. Parameters of this model can be collected in a clue matrix as introduced in section 2.4. The final task is now to recover the actual alignment of words and phrases from the text using the parameters in the clue matrix. This can be formulated as a search task in which one tries to find the optimal alignment using possible links between words and phrases.

It is important for our purposes to allow multiple links from each word (source and target) to corresponding words in the other language in order to obtain phrasal links. We say that a word-to-word link *overlaps* with another one if both of them refer to either the same source or the same target language word. Sets of overlapping links form *link clusters*.

Phrasal links cause alignments with varying numbers of linked items which have to be compared. We use the following dynamic procedure in order to approach an optimal alignment:

1. Find the best link in the clue matrix, i.e. find the word-to-word relation with the highest

value in the matrix. Set the corresponding value in the matrix to zero.

2. Check for overlaps: If the link overlaps with other links from more than one accepted link cluster \rightarrow continue with 1. If the link overlaps with another accepted link but the non-overlapping tokens are not next to each other in the text \rightarrow continue with 1.
3. Add the link to the set of accepted link clusters and continue with 1 until no more links are found (or the best link is below a certain threshold)

The algorithm is very simple and may miss the optimal alignment. However, it is a very efficient way of extracting links according to their association clues. Experiments, which are presented further down, show promising results. The crucial point of the algorithm is the attachment of links to existing link clusters. The algorithm restricts clusters to pairs of contiguous word sequences in order to reduce the number of malformed phrases in extracted links. A better way would be to use proper language models to do this job. Another possibility is to use the syntactic structures from a (shallow) parser as prior knowledge. A simple modification of the algorithm above would be to accept overlapping links only if they do not cross phrase borders according to the syntactic analysis.

4 Bootstrapping Clue Alignment

In section 2.2, we pointed out that clues can be estimated from aligned training material. This allows us to infer new clues from previous links by estimating conditional probabilities. For this, we assume that previous links are correct and can be used for probability estimations. This is not true in general. However, we hope to find additional links with sufficient accuracy from these clues. In other words, we expect clues, which have been found via “self-learning” techniques to increase the recall with an acceptable increase of noise.

Previous links point to the context from which they originated. Therefore, we can access any pair of features which is available for the context as well as for the linked items themselves. In this way, clue probabilities can be based on any combination of features of linked items and their context.

A simple example is to use part-of-speech (POS) tags as a feature of lexical items. Using this feature, we can estimate the probabilities of source language items with certain POS-tags to be linked to target language items with certain other POS-tags.

Consider the following example: An English/Swedish bitext has been aligned with some basic clue patterns using the clue alignment approach. Now, we assume that pairs of POS labels can give us additional clues about possible links. A new clue is, e.g., a conditional probability of a sequence of POS labels of linked items given the POS labels of the items they were linked to.

Applying learned clues (such as the POS clue from above) on their own would probably be misleading in many cases. However, they add valuable information in combination with others. In our experiments, we applied the following feature patterns for learning clues:

- POS full:** Label sequences as described above.
- POS coarse:** Label sequences as in *POS full* but with a reduced tag set for Swedish (done by simply cutting the label after two characters)
- Phrase:** Phrase type labels which have been produced by (shallow) parsers
- Position:** The position of the translations in the target language segment relative to the position of the original in the source language segment.³

The following matrix was produced for the sentence pair from section 2.1 using all the learned alignment clues from above.⁴

	Korridorerna	myllrar	av	dem
The	0.54	0.02	0.17	0.33
corridors	0.82	0.07	0.06	0.06
are	0.17	0.46	0.15	0.14
jumping	0.27	0.73	0.11	0.11
with	0	0	0.63	0.09
them	0	0	0.12	0.61

³The position clue is more of a weight than a clue. It favors common position distances of links by giving them a higher value. This somehow assumes that translations are more likely to be found close to each other than far away in terms of word position.

⁴Each clue has been normalized with a uniform weight of 0.5 except for the position clue which was weighted with a value of 0.1.

The numbers in bold refer to the link clusters which would have been extracted using the clue alignment procedure from section 3.

5 Clue Alignment Experiments

5.1 The Setup

We applied the “clue aligner” to one of our parallel corpora from the PLUG project (Sågval Hein, 1999), a novel by Saul Bellow “To Jerusalem and back: a personal account” with about 170,000 words in English and Swedish.

The English portion of the corpus has been tagged automatically with POS tags by the English maximum entropy tagger in the open-source software package Grok (Baldrige, 2002). The same package was used for shallow parsing of the English sentences.

The Swedish portion was tagged by the Ngram-based TnT-tagger (Brants, 2000) which was trained for Swedish on the SUC corpus (Megyesi, 2001). Furthermore, we used a rule-based analyzer for syntactic parsing (Megyesi, 2002).

Our basic alignment applies two association clues: the Dice coefficient and the longest common subsequence ratio (LCSR). Both clues have been weighted uniformly with a value of 0.5. The threshold for the Dice coefficient has been set to 0.3 and the minimal co-occurrence frequency to 2. The threshold of LCSR scores has been set to 0.4 and the minimal token length to 3 characters.⁵

We certainly wanted to test the ability of finding phrasal links. Therefore, both association clues have been calculated for pairs of multi-word units (MWUs). MWUs may overlap with others or may be included in other MWUs. We used two different approaches in order to select appropriate MWUs:

N-grams: word bigrams and trigrams + simple language filters (stop word lists) to find common phrase borders.

Chunks: phrases which have been marked by a shallow parser for English and a rule-based parser for Swedish.

Both sets of MWUs can also be combined.

⁵Weights and threshold have been chosen intuitively.

Furthermore, we were interested in the ability of the algorithm of learning new clues from previously aligned links as discussed in 4. We applied all the clue patterns which were introduced at the end of section 4: POS full, POS coarse, Phrase, Position. POS clues have been normalized with a weight of 0.5. Relative position and phrase type labels bear much less information about specific words and phrases than POS tags, therefore, a lower weight of 0.1 was chosen for these two clues.

5.2 The results

For the evaluation, we used an existing gold standard which was produced within the PLUG project (Ahrenberg et al., 1999). The gold standard consists of 500 randomly sampled items which have been aligned manually according to detailed guidelines (Merkel, 1999). Results are measured using fine-grained metrics for precision and recall (Ahrenberg et al., 2000) and a balanced F-measure.

The following table summarizes the results of some alignment experiments using different sets of clues (all values in %):⁶

adding clues	precision	recall	F
basic	79.737	41.695	54.757
+ POS full	74.311	49.554	59.458
+ POS coarse	69.854	57.279	62.945
+ phrase	78.289	45.122	57.249
+ position	81.939	44.703	57.847
+ all	74.749	63.730	68.801

The values in the table above show a clear improvement, mainly in recall, of the results when additional clues have been applied. The impact of POS clues follows intuitive expectations. A smaller tagset makes the clues more general and the recall value is increased while the precision drops significantly. However, the position clue changed the result beyond all expectations. Not only did recall go up significantly, even the precision was increased. This proves that relations between word positions are important in aligning Swedish and English. Position distance weights (implicitly implemented as a position clue) seem to improve the choice between competing link alternatives. The effect of similar clues on less re-

⁶The values refer to experiments with the “chunk” approach for MWU selection.

lated languages is necessary in order to evaluate their general quality.

The impact of the MWU approach was investigated as well. The following table compares results of the two different approaches as well as the combined approach.

MWU approach	precision	recall	F
N-grams	74.462	62.539	67.981
chunks	74.749	63.730	68.801
N-grams + chunks	70.894	61.115	65.642

The “chunk” approach is best in all categories. The lower values for precision for the other two approaches could be expected. However, the low recall value for the combined approach is a surprise.

6 Conclusions

In this paper, a word alignment approach has been introduced which is based on a combination of association clues. The algorithm supports a dynamic tokenization of parallel texts which enables the alignment system to combine relations of overlapping pairs of lexical items (words and phrases).

Alignment clues can be estimated from common association criteria such as co-occurrence statistics and string similarity measures. Clues can also be learned from pre-aligned training data. It has been demonstrated how self-learning techniques can be used for learning additional alignment clues from previous alignments. Clues are defined as probabilities which indicate an association between lexical items according to some of their features. This definition is very flexible as features can represent any kind of information that is available for each item and its context. An important advantage of the clue alignment algorithm is the possibility of combining association scores. In this way, any number of clues can be included in the alignment process.

The alignment experiments, which were presented in this paper, demonstrate the combination of two basic clue types (based on co-occurrence and string similarity) with four additional clue types (based on POS labels, chunk labels, and relative word positions) which were learned during the alignment process. The alignment experiments on a parallel English/Swedish corpus showed significant improvements of the results when additional

clues were added to the basic settings.

The clue alignment approach is very flexible and can easily be adapted to a new domain, language pair, and a different set of clues. Clue patterns can be defined depending on the information which is available (POS tags, phrase types, semantic tags, named entity markup, dictionaries etc.). However, clue patterns have to be designed carefully as they can be misleading. Word alignment is a real-life problem: We are looking for links in the complex world of parallel corpora and we need good clues in order to find them.

References

- Lars Ahrenberg, Magnus Merkel, and Mikael Andersson. 1998. A simple hybrid aligner for generating lexical correspondences in parallel texts. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics*, pages 29–35, Montreal, Canada.
- Lars Ahrenberg, Magnus Merkel, Anna Sgvall Hein, and Jrg Tiedemann. 1999. Evaluation of LWA and UWA. Technical Report 15, Department of Linguistics, University of Uppsala.
- Lars Ahrenberg, Magnus Merkel, Anna Sgvall Hein, and Jrg Tiedemann. 2000. Evaluation of word alignment systems. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation, LREC-2000*. European Language Resources Association.
- Lars Ahrenberg, Magnus Merkel, and Mikael Andersson. 2002. A system for incremental and interactive word linking. In *Proceedings from The Third International Conference on Language Resources and Evaluation (LREC-2002)*, pages 485–490, Las Palmas, Spain.
- Jason Baldridge. 2002. GROK - an open source natural language processing library. <http://grok.sourceforge.net/>.
- Thorsten Brants. 2000. TnT - a statistical part-of-speech tagger. In *Proceedings of the Sixth Applied Natural Language Processing Conference ANLP-2000*, Seattle, Washington.
- Djoerd Hiemstra. 1998. Multilingual domain modeling in Twenty-One: Automatic creation of a bidirectional translation lexicon from a parallel corpus. In Peter-Arno Coppen, Hans van Halteren, and Lisanne Teunissen, editors, *Proceedings of the eighth CLIN meeting*, pages 41–58.
- Beta Megyesi. 2001. Comparing data-driven learning algorithms for POS tagging of swedish. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, pages 151–158, Carnegie Mellon University, Pittsburgh, PA, USA.
- Beta Megyesi. 2002. Shallow parsing with POS taggers and linguistic features. *Journal of Machine Learning Research: Special Issue on Shallow Parsing*, (2):639–668.
- I. Dan Melamed. 1995. Automatic evaluation and uniform filter cascades for inducing N-best translation lexicons. In *Proceedings of the 3rd Workshop on Very Large Corpora*, Boston/Massachusetts.
- I. Dan Melamed. 1997. Automatic discovery of non-compositional compounds in parallel data. In *Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing*, Providence.
- I. Dan Melamed. 1998. Annotation style guide for the Blinker project, version 1.0. IRCS Technical Report 98-06, University of Pennsylvania, Philadelphia.
- Magnus Merkel. 1999. Annotation style guide for the PLUG link annotator. Technical report, Linkping University, Linkping.
- Franz Josef Och and Hermann Ney. 2000. Improved statistical alignment models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447.
- Anna Sgvall Hein. 1999. The PLUG Project: Parallel corpora in Linkping, Uppsala, and Gteborg: Aims and achievements. Technical Report 16, Department of Linguistics, University of Uppsala.
- Frank Smadja, Kathleen R. McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, 22(1), pages 1–38.
- Jrg Tiedemann. 1999. Word alignment - step by step. In *Proceedings of the 12th Nordic Conference on Computational Linguistics NODALIDA99*, pages 216–227, University of Trondheim, Norway.
- Dan Tufis and Ana-Maria Barbu. 2002. Lexical token alignment: Experiments, results and applications. In *Proceedings from The Third International Conference on Language Resources and Evaluation (LREC-2002)*, pages 458–465, Las Palmas, Spain.
- Stephan Vogel, Franz Josef Och, Christoph Tillmann, Sonja Nieen, Hassan Sawaf, and Hermann Ney, 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*, chapter Statistical Methods for Machine Translation. Springer Verlag, Berlin.