

Word to word alignment strategies

Jörg Tiedemann

Department of Linguistics and Philology
Uppsala University
Uppsala/Sweden
joerg@stp.ling.uu.se

Abstract

Word alignment is a challenging task aiming at the identification of translational relations between words and multi-word units in parallel corpora. Many alignment strategies are based on links between single words. Different strategies can be used to find the optimal word alignment using such one-to-one word links including relations between multi-word units. In this paper seven algorithms are compared using a word alignment approach based on association clues and an English-Swedish bitext together with a handcrafted reference alignment used for evaluation.

1 Introduction

Word alignment is the task of identifying translational relations between words in parallel corpora with the aim of re-using them in natural language processing. Typical applications that make use of word alignment techniques are machine translation and multi-lingual lexicography. Several approaches have been proposed for the automatic alignment of words and phrases using statistical techniques and alignment heuristics, e.g. (Brown et al., 1993; Vogel et al., 1996; García-Varea et al., 2002; Ahrenberg et al., 1998; Tiedemann, 1999; Tufis and Barbu, 2002; Melamed, 2000). Word alignment usually includes links between so-called multi-word units (MWUs) in cases where lexical items cannot be split into separated words with appropriate translations in another language. See for example the alignment between an English sentence and a Swedish sentence illustrated in figure 1. There are MWUs in both languages aligned to corresponding translations in the other language. The Swedish compound “mittplatsen” corresponds to three words in English (“the middle seat”) and the English verb “dislike” is translated into a Swedish particle verb “tycker om” (English: like) that has been negated using “inte”. Most approaches model

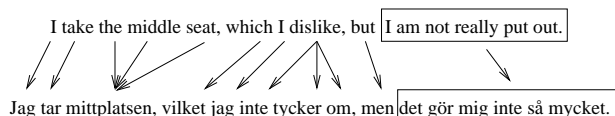


Figure 1: A word alignment example from Saul Bellow “To Jerusalem and back: a personal account” (Bellow, 1976) and its Swedish translation (Bellow, 1977) (the *Bellow corpus*).

word alignment as links between words in the source language and words in the target language as indicated by the arrows in figure 1. However, in cases like the English expression “I am not really put out” which corresponds to the Swedish expression “det gör mig inte så mycket” there is no proper way of connecting single words with each other in order to express this relation. In some approaches such relations are constructed in form of an exhaustive set of links between all word pairs included in both expressions (Melamed, 1998; Mihalcea and Pedersen, 2003). In other approaches complex expressions are identified in a pre-processing step in order to handle them as complex units in the same manner as single words in alignment (Smadja et al., 1996; Ahrenberg et al., 1998; Tiedemann, 1999).

The one-to-one word linking approach seems to be very limited. However, single word links can be combined in order to describe links between multi-word units as illustrated in figure 1. In this paper we investigate different alignment strategies using this approach¹. For this we apply clue alignment introduced in the next section.

2 Word alignment with clues

The clue alignment approach has been presented in (Tiedemann, 2003). Alignment clues represent probabilistic indications of associa-

¹A similar study on statistical alignment models is included in (Och and Ney, 2003).

tions between lexical items collected from different sources. *Declarative clues* can be taken from linguistic resources such as bilingual dictionaries. They may also include pre-defined relations between lexical items based on certain features such as parts of speech. *Estimated clues* are derived from the parallel data using, for example, measures of co-occurrence (e.g. the Dice coefficient (Smadja et al., 1996)), statistical alignment models (e.g. IBM models from statistical machine translation (Brown et al., 1993)), or string similarity measures (e.g. the longest common sub-sequence ratio (Melamed, 1995)). They can also be learned from previously aligned training data using linguistic and contextual features associated with aligned items. Relations between certain word classes with respect to the translational association of words belonging to these classes is one example of such clues that can be learned from aligned training data. In our experiments, for example, we will use clues that indicate relations between lexical items based on their part-of-speech tags and their positions in the sentence relative to each other. They are learned from automatically word-aligned training data.

The clue alignment approach implements a way of combining association indicators on a word-to-word level. The combination of clues results in a two-dimensional clue matrix. The values in this matrix express the collected evidence of an association between word pairs in bitext segments taken from a parallel corpus. Word alignment is then the task of identifying the best links according to the associations indicated in the clue matrix. Several strategies for such an alignment are discussed in the following section.

3 Alignment strategies

A clue matrix summarizes information from various sources that can be used for the identification of translation relations. However, there is no obvious way to utilize this information for word alignment as we explicitly include multi-word units (MWUs) in our approach. The clue matrix in figure 2 has been obtained for a bitext segment from our English-Swedish test corpus (the Bellow corpus) using a set of weighted declarative and estimated clues.

There are many ways of “clustering” words together and there is no obvious maximization procedure for finding the alignment optimum when MWUs are involved. The alignment pro-

	ingen	visar	särskilt	mycket	tålmod
no	29	0	0	1	9
one	16	2	1	1	13
is	1	13	1	2	0
very	0	2	18	17	1
patient	2	1	4	12	6

Figure 2: A clue matrix (all values in %).

cedure depends very much on the definition of an optimal alignment. The best alignment for our example would probably be the set of the following links:

$$\text{links} = \left\{ \begin{array}{ll} \text{no one} & \text{ingen} \\ \text{is patient} & \text{visar tålmod} \\ \text{very} & \text{särskilt mycket} \end{array} \right\}$$

A typical procedure for automatic word alignment is to start with one-to-one word links. Links that have common source or target language words are called *overlapping links*. Sets of overlapping links, which do not overlap with any other link outside the set, are called *link clusters (LC)*. Aligning words one by one often produces overlaps and in this way implicitly creates aligned multi-word-units as part of link clusters. A general word-to-word alignment L for a given bitext segment with N source language words ($s_1 s_2 \dots s_N$) and M target language words ($t_1 t_2 \dots t_M$) can be formally described as a set of links $L = \{L_1, L_2, \dots, L_x\}$ with $L_x = [s_{x_1}, t_{x_2}]$, $x_1 \in \{1..N\}$, $x_2 \in \{1..M\}$.

This general definition allows varying numbers of links ($0 \leq x \leq N * M$) within possible alignments L . It is not straightforward how to find the optimal alignment as L may include different numbers of links.

3.1 Directional alignment models

One word-to-word alignment approach is to assume a *directional word alignment model* similar to the models in statistical machine translation. The directional alignment model assumes that there is at most one link for each source language word. Using alignment clues, this can be expressed as the following optimization problem: $\hat{L}^D = \text{argmax}_{L^D} \prod_{n=1}^N C(L_n^D)$ where $L^D = \{L_1^D, L_2^D, \dots, L_N^D\}$ is a set of links $L_n^D = [s_n, t_{a_n^D}]$ with $a_n^D \in \{1..M\}$ and $C(L_n^D)$ is the combined clue value for the linked items s_n and $t_{a_n^D}$. In other words, word alignment is the search for the best link for each source language word. Directional models do not allow multiple links from one item to several target items. However, target items can be linked

to multiple source language words as they can be aligned to the same target language word. The direction of alignment can easily be reversed, which leads to the *inverse directional alignment*: $\hat{L}^I = \operatorname{argmax}_{L^I} \prod_{m=1}^M C(L_m^I)$ with links $L_m^I = [s_{a_m^I}, t_m]$ and $a_m^I \in \{1..N\}$. In the inverse directional alignment, source language words can be linked to multiple words but not the other way around. The following figure illustrates directional alignment models applied to the example in figure 2:

$$\hat{L}^D = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{one} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \\ \text{patient} & \text{mycket} \end{pmatrix}$$

$$LC^D = \begin{pmatrix} \text{no one} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \\ \text{patient} & \text{mycket} \end{pmatrix}$$

Using the inverse directional alignment strategy we would obtain the following links:

$$\hat{L}^I = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \\ \text{very} & \text{mycket} \\ \text{one} & \text{tålåmod} \end{pmatrix}$$

$$LC^I = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt mycket} \\ \text{one} & \text{tålåmod} \end{pmatrix}$$

3.2 Combined directional alignment

Directional link sets can be combined in several ways. The *union* of link sets ($\hat{L}^U = \hat{L}^D \cup \hat{L}^I$) usually causes many overlaps and, hence, very large link clusters. On the other hand, an *intersection* of link sets ($\hat{L}^\cap = \hat{L}^D \cap \hat{L}^I$) removes all overlaps and leaves only highly confident one-to-one word links behind. Using the same example from above we obtain the following alignments:

$$\hat{L}^U = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{one} & \text{ingen} \\ \text{one} & \text{tålåmod} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \\ \text{very} & \text{mycket} \\ \text{patient} & \text{mycket} \end{pmatrix}$$

$$LC^U = \begin{pmatrix} \text{no one} & \text{ingen} & \text{tålåmod} \\ \text{is} & \text{visar} & \\ \text{very patient} & \text{särskilt} & \text{mycket} \end{pmatrix}$$

The intersection of links produces the following sets:

$$\hat{L}^\cap = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \end{pmatrix}$$

$$LC^\cap = \hat{L}^\cap$$

The union and the intersection of links do not produce satisfactory results as seen in the example. Another alignment strategy is a refined combination of link sets ($\hat{L}^R = \{\hat{L}^D \cap \hat{L}^I\} \cup \{L_1^R, \dots, L_r^R\}$) as suggested by (Och and Ney, 2000b). In this approach, the intersection of links is iteratively extended by additional links L_r^R which pass **one** of the following two constraints:

- A new link is accepted if both items in the link are not yet aligned.
- Mapped on a two-dimensional bitext space, the new link is either vertically or horizontally adjacent to an existing link **and** the new link does not cause any link to be adjacent to other links in both dimensions (horizontally and vertically).

Applying this approach to the example, we get:

$$\hat{L}^R = \begin{pmatrix} \text{no} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt} \\ \text{very} & \text{mycket} \\ \text{one} & \text{ingen} \\ \text{patient} & \text{tålåmod} \end{pmatrix}$$

$$LC^R = \begin{pmatrix} \text{no one} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very} & \text{särskilt mycket} \\ \text{patient} & \text{tålåmod} \end{pmatrix}$$

3.3 Competitive linking

Another alignment approach is the *competitive linking approach* proposed by Melamed (Melamed, 1996). In this approach, one assumes that there are **only** one-to-one word links. The alignment is done in a greedy “best-first” search manner where links with the highest association scores are aligned first, and the aligned items are then immediately removed from the search space. This process is repeated until no more links can be found. In this way, the optimal alignment (\hat{L}^C) for non-overlapping one-to-one links is found. The number of possible links in an alignment is reduced to $\min(N, M)$. Using competitive linking with our example we yield:

$$\hat{L}^C = \left\{ \begin{array}{ll} \text{no} & \text{ingen} \\ \text{very} & \text{särskilt} \\ \text{is} & \text{visar} \\ \text{one} & \text{tålåmod} \\ \text{patient} & \text{mycket} \end{array} \right\}$$

$$LC^C = \hat{L}^C$$

3.4 Constrained best-first alignment

Another iterative alignment approach has been proposed in (Tiedemann, 2003). In this approach, the link $L_x^B = [s_{x_1}, t_{x_2}]$ with the highest score in the clue matrix $\hat{C}(s_{x_1}, t_{x_2}) = \max_{s_i, t_j}(C(s_i, t_j))$ is added to the set of link clusters if it fulfills certain constraints. The top score is removed from the matrix (i.e. set to zero) and the link search is repeated until no more links can be found. This is basically a *constrained best-first search*. Several constraints are possible. In (Tiedemann, 2003) an adjacency check is suggested, i.e. overlapping links are accepted only if they are adjacent to other links in one and only one existing link cluster. Non-overlapping links are always accepted (i.e. a non-overlapping link creates a new link cluster). Other possible constraints are clue value thresholds, thresholds for clue score differences between adjacent links, or syntactic constraints (e.g. that link clusters may not cross phrase boundaries). Using a best-first search strategy with the adjacency constraint we obtain the following alignment:

$$\hat{L}^B = \left\{ \begin{array}{ll} \text{no} & \text{ingen} \\ \text{very} & \text{särskilt} \\ \text{very} & \text{mycket} \\ \text{one} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{patient} & \text{mycket} \\ \text{patient} & \text{tålåmod} \end{array} \right\}$$

$$LC^B = \left\{ \begin{array}{ll} \text{no one} & \text{ingen} \\ \text{is} & \text{visar} \\ \text{very patient} & \text{särskilt mycket tålåmod} \end{array} \right\}$$

3.5 Summary

None of the alignment approaches described above produces the preferred reference alignment in our example using the given clue matrix. However, simple iterative procedures come very close to the reference and produce acceptable alignments even for multi-word units, which is promising for an automatic clue alignment system. Directional alignment models depend very much on the relation between the

source and the target language. One direction usually works better than the other, e.g. an alignment from English to Swedish is better than Swedish to English because in English terms and concepts are often split into several words whereas Swedish tends to contain many compositional compounds. Symmetric approaches to word alignment are certainly more appropriate for general alignment systems than directional ones.

4 Evaluation methodology

Word alignment quality is usually measured in terms of precision and recall. Often, previously created *gold standards* are used as reference data in order to simplify automatic tests of alignment attempts. Gold standards can be re-used for additional test runs which is important when examining different parameter settings. However, recall and precision derived from information retrieval have to be adjusted for the task of word alignment. The main difficulty with these measures in connection with word alignment arises with links between MWUs that cause *partially correct* alignments. It is not straightforward how to judge such links in order to compute precision and recall. In order to account for partiality we use a slightly modified version of the partiality score Q proposed in (Ahrenberg et al., 2000)²:

$$Q_x^{precision} = \frac{|alg_{src}^x \cap corr_{src}^x| + |alg_{trg}^x \cap corr_{trg}^x|}{|alg_{src}^x| + |alg_{trg}^x|}$$

$$Q_x^{recall} = \frac{|alg_{src}^x \cap corr_{src}^x| + |alg_{trg}^x \cap corr_{trg}^x|}{|corr_{src}^x| + |corr_{trg}^x|}$$

The set of alg_{src}^x includes all source language words of **all** proposed links if at least one of them is partially correct with respect to the reference link x from the gold standard. Similarly, alg_{trg}^x refers to all the proposed target language words. $corr_{src}^x$ and $corr_{trg}^x$ refer to the sets of source and target language words in link x of the gold standard. Using the partiality value Q , we can define the recall and precision metrics as follows:

$$R_{mwu} = \frac{\sum_{x=1}^X Q_x^{recall}}{|correct|}, P_{mwu} = \frac{\sum_{x=1}^X Q_x^{precision}}{|aligned|}$$

A balanced F-score can be used to combine both, precision and recall:

$$F_{mwu} = (2 * P_{mwu} * R_{mwu}) / (P_{mwu} + R_{mwu}).$$

² $Q_x \equiv 0$ for incorrect links for both, precision and recall.

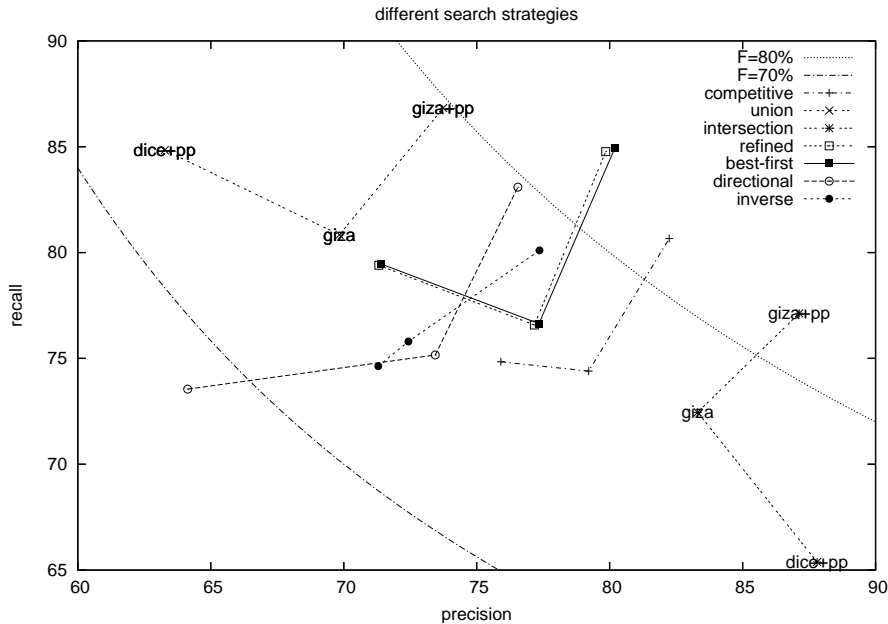


Figure 3: Different alignment search strategies. Clue alignment settings: *dice+pp*, *giza*, and *giza+pp*. Alignment strategies: *directional* (L^D), *inverse directional* (L^I), *union* (L^U), *intersection* (L^\cap), *refined* (L^R), *competitive linking* (L^C), and constrained *best-first* (L^B).

Alternative measures for the evaluation of one-to-one word links have been proposed in (Och and Ney, 2000a; Och and Ney, 2003). However, these measures require completely aligned bitext segments as reference data. Our gold standards include random samples from the corpus instead (Ahrenberg et al., 2000). Furthermore, we do not split MWU links as proposed by (Och and Ney, 2000a). Therefore, the measures proposed above are a natural choice for our evaluations.

5 Experiments

Several alignment search strategies have been discussed in the previous sections. Our clue aligner implements these strategies in order to test their impact on the alignment performance. In the experiments we used one of our English-Swedish bitext from the PLUG corpus (Sågwall Hein, 2002), the novel “To Jerusalem and back: A personal account” by Saul Bellow. This corpus is fairly small (about 170,000 words) and therefore well suited for extensive studies of alignment parameters. For evaluation, a gold standard of 468 manually aligned links is used (Merkel et al., 2002). It includes 122 links with MWUs either on the source or on the target side (= 26% of the gold standard). 109 links contain source language MWUs, 59 links target language MWUs, and 46 links MWUs in both

languages. 10 links are *null* links, i.e. a link of one word to an empty string. Three different clue types are used for the alignment: the Dice coefficient (*dice*), lexical translation probabilities derived from statistical translation models (*giza*) using the GIZA++ toolbox (Och and Ney, 2003), and, finally, POS/relative-word-position-clues learned from previous alignments (*pp*). Alignment strategies are compared on the basis of three different settings: *dice+pp*, *giza*, and *giza+pp*. In figure 3, the alignment results are shown for the three clue settings using different search strategies as discussed earlier.

5.1 Discussion

Figure 3 illustrates the relation between precision and recall when applying different algorithms. As expected, the intersection of directional alignment strategies yields the highest precision at the expense of recall, which is generally lower than for the other approaches. Contrary to the intersection, the union of directional links produces alignments with the highest recall values but lower precision than all other search algorithms. Too many (partially) incorrect MWUs are included in the union of directional links. The intersection on the other hand includes only one-to-one word links that tend to be correct. However, many links are missed in this strategy evident in the low recall val-

<i>giza+pp</i>	non-MWU		MWU-links (122 in total)				English MWU		Swedish MWU		both	
	P	R	correct	partial	P	R	P	R	P	R	P	R
directional	82.66	88.73	19	77	59.23	67.10	58.04	69.84	68.93	70.16	68.85	77.52
inverse	81.93	87.28	5	50	64.39	59.74	64.68	57.87	67.62	72.57	69.21	71.78
union	80.13	91.62	21	88	55.61	73.24	55.57	73.29	63.40	81.74	65.50	84.25
intersection	91.67	85.84	0	98	74.35	52.44	73.56	53.44	83.04	60.31	83.33	64.89
competitive	88.44	88.44	0	105	64.66	58.59	66.11	59.71	72.13	67.94	77.66	73.23
refined	84.61	88.15	28	78	65.91	72.61	65.53	72.70	74.37	80.56	75.86	83.03
best-first	85.07	89.02	28	79	66.40	73.40	66.23	73.77	75.38	81.35	77.52	84.48

Table 1: Evaluations of different link types for the setting *giza+pp*.

ues. Directional alignment strategies generally yield lower F-values than other refined symmetric alignment strategies. Their implementation is straightforward but the results are highly dependent on the language pair under consideration. The differences between the two alignment directions in our example are surprisingly inconsistent. Using the *giza* clues both alignment results are very close in terms of precision and recall whereas a larger difference can be observed using the other two clue settings when applying different directional alignment strategies. Competitive linking is somewhat in between the intersection approach and the two symmetric approaches, “best-first” and “refined”. This could also be expected as competitive linking only allows non-overlapping one-to-one word links. The refined bi-directional alignment approach and the constrained best-first approach are almost identical in our examples with a more or less balanced relation between precision and recall. One advantage of the best-first approach is the possibility of incorporating different constraints that suit the current task. The adjacency check is just one of the possible constraints. For example, syntactic criteria could be applied in order to force linked items to be complete according to existing syntactic markup. Non-contiguous elements could also be identified using the same approach simply by removing the adjacency constraint. However, this seems to increase the noise significantly according to experiments not shown in this paper. Further investigations on optimizing alignment constraints for certain tasks have to be done in the future. Focusing on MWUs, the numbers in table 1 show a clear picture about the difficulties of all approaches to find correct MWU links. Symmetric alignment strategies like *refined* and *best-first* produce in general the best results for MWU links. However, the main portion of such links is only partially correct even for these approaches. Using our partiality measure, the

intersection of directional alignments still produces the highest precision values when considering MWU links only even though no MWUs are included in these alignments at all. The best results among MWU links are achieved for the ones including MWUs in both languages. However, these results are still significantly lower than for single-word links (*non-MWU*).

6 Conclusions

According to our results different alignment strategies can be chosen to suit particular needs. Concluding from the experiments, restrictive methods like the intersection of directional alignments or competitive linking should be chosen if results with high precision are required (which are mostly found among one-to-one word links). This is, for example, the case in automatic extraction of bilingual lexicons where noise should be avoided as much as possible. A strong disadvantage of these approaches is that they do not include MWUs at all. Other strategies should be chosen for applications, which require a comprehensive coverage as, for example, machine translation. Symmetric approaches such as the refined combination of directional alignments and the constrained best-first alignment strategy yield the highest overall performance. They produce the best balance between precision and recall and the highest scores in terms of F-values.

References

- Lars Ahrenberg, Magnus Merkel, and Mikael Andersson. 1998. A simple hybrid aligner for generating lexical correspondences in parallel texts. In Christian Boitet and Pete White-lock, editors, *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agvall Hein, and Jorg Tiedemann. 2000. Evaluation

- of word alignment systems. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, volume III, pages 1255–1261, Athens, Greece.
- Saul Bellow. 1976. *From Jerusalem and back: a personal account*. The Viking Press, New York, USA.
- Saul Bellow. 1977. *Jerusalem tur och retur*. Bonniers, Stockholm. Translation of Caj Lundgren.
- Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311, June.
- Ismael García-Varea, Franz Josef Och, Hermann Ney, and Francisco Casacuberta. 2002. Improving alignment quality in statistical machine translation using context-dependent maximum entropy models. In *Proceedings of the 19th International Conference on Computational Linguistics*, pages 1051–1054, Taipei, Taiwan, August.
- I. Dan Melamed. 1995. Automatic evaluation and uniform filter cascades for inducing n-best translation lexicons. In David Yarovsky and Kenneth Church, editors, *Proceedings of the 3rd Workshop on Very Large Corpora*, pages 184–198, Boston, MA. Association for Computational Linguistics.
- I. Dan Melamed. 1996. Automatic construction of clean broad-coverage lexicons. In *Proceedings of the 2nd Conference the Association for Machine Translation in the Americas*, pages 125–134, Montreal, Canada.
- I. Dan Melamed. 1998. Annotation style guide for the Blinker project, version 1.0. IRCS Technical Report 98-06, University of Pennsylvania, Philadelphia, PA.
- I. Dan Melamed. 2000. Models of translational equivalence among words. *Computational Linguistics*, 26(2):221–249, June.
- Magnus Merkel, Mikael Andersson, and Lars Ahrenberg. 2002. The PLUG link annotator - interactive construction of data from parallel corpora. In Lars Borin, editor, *Parallel Corpora, Parallel Worlds*. Rodopi, Amsterdam, New York. Proceedings of the Symposium on Parallel Corpora, Department of Linguistics, Uppsala University, Sweden, 1999.
- Rada Mihalcea and Ted Pedersen. 2003. An evaluation exercise for word alignment. In *Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pages 1–10, Edmonton, Canada, May.
- Franz-Josef Och and Hermann Ney. 2000a. A comparison of alignment models for statistical machine translation. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 1086–1090, Saarbrücken, Germany, July.
- Franz Josef Och and Hermann Ney. 2000b. Improved statistical alignment models. In *Proc. of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 440–447.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- Anna Sågvall Hein. 2002. The PLUG project: Parallel corpora in Linköping, Uppsala, and Göteborg: Aims and achievements. In Lars Borin, editor, *Parallel Corpora, Parallel Worlds*. Rodopi, Amsterdam, New York. Proceedings of the Symposium on Parallel Corpora, Department of Linguistics, Uppsala University, Sweden, 1999.
- Frank A. Smadja, Kathleen R. McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, 22(1), pages 1–38.
- Jörg Tiedemann. 1999. Word alignment - step by step. In *Proceedings of the 12th Nordic Conference on Computational Linguistics*, pages 216–227, University of Trondheim, Norway.
- Jörg Tiedemann. 2003. Combining clues for word alignment. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 339–346, Budapest, Hungary, April.
- Dan Tufis and Ana-Maria Barbu. 2002. Lexical token alignment: Experiments, results and applications. In *Proceedings from The 3rd International Conference on Language Resources and Evaluation*, pages 458–465, Las Palmas, Spain.
- Stephan Vogel, Hermann Ney, and Christoph Tillmann. 1996. HMM-based word alignment in statistical translation. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 836–841, Copenhagen, Denmark.