ATLAS – a new text alignment architecture

Bettina Schrader
Institute of cognitive Science
University of Osnabrück
49069 Osnabrück
bschrade@uos.de

Abstract

We are presenting a new, hybrid alignment architecture for aligning bilingual, linguistically annotated parallel corpora. It is able to align simultaneously at paragraph, sentence, phrase and word level, using statistical and heuristic cues, along with linguistics-based rules. The system currently aligns English and German texts, and the linguistic annotation used covers POS-tags, lemmas and syntactic constitutents. However, as the system is highly modular, we can easily adapt it to new language pairs and other types of annotation.

The hybrid nature of the system allows experiments with a variety of alignment cues to find solutions to word alignment problems like the correct alignment of rare words and multiwords, or how to align despite syntactic differences between two languages.

First performance tests are promising, and we are setting up a gold standard for a thorough evaluation of the system.

1 Introduction

Aligning parallel text, i.e. automatically setting the sentences or words in one text into correspondence with their equivalents in a translation, is a very useful preprocessing step for a range of applications, including but not limited to machine translation (Brown et al., 1993), cross-language information retrieval (Hiemstra, 1996), dictionary creation (Smadja et al., 1996) and induction of NLP-tools (Kuhn, 2004). Aligned corpora can be also be used in translation studies (Neumann and Hansen-Schirra, 2005).

The alignment of sentences can be done sufficiently well using cues such as sentence length (Gale and Church, 1993) or cognates (Simard et al., 1992). Word alignment, however, is almost exclusively done using statistics (Brown et al., 1993; Hiemstra, 1996; Vogel et al., 1999; Toutanova et al., 2002).

Hence it is difficult to align so-called rare events, i.e. tokens with a frequency below 10. This is a considerable drawback, as rare events make up more than half of the vocabulary of any corpus. Another problem is the correct alignment of multiword units like idioms. Then, differences in word order are not modelled well by the statistical algorithms.

In order to find solutions to these problems, we have developed a hybrid alignment architecture: it uses statistical information extracted directly from a corpus, and rules or heuristics based on the linguistic information as given by the corpus' annotation. Additionally, it is not necessary to compute sentence alignment prior to aligning at the word level. Instead, the system is capable of interactively and incrementally computing sentence and word alignment, along with alignment at the paragraph and phrase level. The simultaneous alignment at different levels of granularity imposes restrictions on the way text alignment is computed: we are using a constrained best-first strategy for this purpose.

Although we are currently developing and testing the alignment system for the language pair English-German, we have made sure that it can easily be extended to new language pairs. In fact, we are currently adding Swedish and French to the set of supported languages.

First performance tests have been promising, and we are currently setting up a gold standard

Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions, pages 715–722,
Sydney, July 2006.

of 242 manually aligned sentence pairs in English and German for a thorough evaluation.

In the following, we give an overview on standard approaches to sentence and word alignment, and discuss their advantages and shortcomings. Then, we describe the design of our alignment architecture. In the next two sections, we are describing the data on which we test our system, and our evaluation strategy. Finally, we sum up and describe further work.

2 Related work

Research on text alignment has largely focused on aligning either sentences or words, i.e. most approaches either compute which sentences of a source and a target language form a translation pair, or they use sentence alignment as a preprocessing step to align on the word level.

Additionally, emphasis was laid on the development of *language-independent* algorithms. Ideally, such algorithms would not be tailored to align a specific language pair, but would be applicable to any two languages. Language-independence has also been favoured with respect to linguistic resources in that alignment should do without e.g. using pre-existing dictionaries. Hence there is a dominance of purely statistical approaches.

2.1 Sentence Alignment

Sentence alignment strategies fall roughly into three categories: length-based approaches (Gale and Church, 1991; Gale and Church, 1993) are based on the assumption that the length proportions of a sentence and its translation are roughly the same. Anchor-based algorithms align sentences based on cues like corpus-specific markup and orthographic similarity (Simard et al., 1992). The third approach uses bilingual lexical information, e.g. estimated from the corpus (Kay and Röscheisen, 1993; Fung and Church, 1994; Fung and McKeown, 1994).

Hybrid methods (Tschorn, 2002) combine these standard approaches such that the shortcomings of one approach are counterbalanced by the strength of another component: length-based methods are very sensitive towards deletions in that a single omission can cause the alignment to go on a wrong track from the point where it occurred to the end of the corpus. Strategies that assume that orthographic similarity entails translational equivalence rely on the relatedness of the language pair in question. In closely-related languages like English and French, the amount of orthographically similar words that share the same meaning is higher than in unrelated languages like English and Chinese, were orthographic or even phonetic similarity may only indicate translational equivalence for names. Strategies that use system-external dictionaries, finally, can only be used if a large-enough dictionary exists for a specific language pair.

2.2 Word Alignment

Aligning below the sentence level is usually done using statistical models for machine translation (Brown et al., 1991; Brown et al., 1993; Hiemstra, 1996; Vogel et al., 1999) where any word of the target language is taken to be a possible translation for each source language word. The probability of some target language word to be a translation of a source language word then depends on the frequency with which both co-occur at the same or similar positions in the parallel corpus.

The probabilities are estimated from the using the EM-algorithm[1], and a Viterbi search is carried out to compute the most probable sequence of word translation pairs. Word order differences between the two languages are modelled by using statistical weights, and multiword units are similarly treated.

Another approach to word alignment is presented by Tiedemann (2003), where alignment probabilities are computed using a combination of features like e.g. co-occurrence, cognateness, syntactic category membership. However, although the alignment is partly based on linguistic features, its computation is entirely statistical. Other word alignment strategies (Toutanova et al., 2002; Cherry and Lin, 2003) have also begun to incorporate linguistic knowledge. Unfortunately, the basic, statistical, assumptions have not been changed, and hence no sufficient solution to the shortcomings of the early alignment models have been found.

3 Shortcomings of the statistical alignment approaches

While sentence alignment can be done successfully using a combination of the existing algorithms, word alignment quality suffers due to three problematic phenomena: the amount of *rare*

[1] see (Manning and Schütze, 1999), chapter 14.2.2 for a general introduction

words typically found in corpora, *word order differences* between the to languages to be aligned, and the existence of *multiword units*

3.1 Rare Words

Approximately half of a corpus' vocabulary consists of so-called *hapax legomena*, i.e. types that occur exactly once in a text. Most other words fall into the range of so-called *rare events*, which we define here as types with occurrences between 2 and 10. Both hapax legomena and rare events obviously do not provide sufficient information for statistical analysis.

In the case of word alignment, it is easy to see that they are hard to align: there is virtually no frequency or co-occurrence data with which to compute the alignment. On the other hand, five to ten percent of a corpus' vocabulary consists of highly frequent words, i.e. words with frequencies of 100 or above. These types have the advantage of occurring frequently enough for statistical analysis, however, as they occur at virtually every position in a corpus, they can correspond to anything if alignment decisions are taken on the basis of statistics only.

One solution to this problem would be to use statistics-free rules for alignment, i.e. rules that are insensitive to the rarity or frequency of a word. However, this means that statistical models either have to be abandoned completely, or that effort has to be put in finding a means to combine both alignment approaches into one single, hybrid system.

An alternative would be to design a statistical alignment model that is better suited for the Zipfian frequency distributions in the source and target language texts. Research in this direction would greatly benefit from large amounts of high quality example alignments, e.g. taken from the parallel treebanks that are currently being built (Volk and Samuelsson, 2004; Neumann and Hansen-Schirra, 2005).

3.2 Word Order Differences

Another problem that has been noticed as early as 1993 with the first research on word alignment (Brown et al., 1993) concerns the differences in word order between source and target language.

While simple statistical alignment models like IBM-1 (Brown et al., 1993) and the symmetric alignment approach by Hiemstra (1996) treat sentences as unstructured bags of words, the more sophisticated IBM-models by Brown et al. (1993) approximates word order differences using a statistical *distortion* factor. Vogel et al. (1999), on the other hand, treat word order differences as a local phenomenon that can be modelled within a window of no more than three words. Recently, researchers like Cherry and Lin (2003) have begun to use syntactic analyses to guide and restrict the word alignment process.

The advantage of using available syntactic information for word alignment is that it helps to overcome data sparseness: although a token may be rare, its syntactic category may not, and hence there may be sufficient statistical information to align at the phrase level. Subsequently, the phrase level information can be used to compute alignments for the tokens within the aligned phrases. The syntactic function of a token as modifier, head etc. can equally simplify and guide the alignment process considerably. However, it is unclear whether such an approach performs well for language pairs where syntactic and functional differences are greater than between e.g. English and French.

3.3 Multiword alignment

Like syntactic differences, n:m correspondences, i.e. alignments that involve multiword expressions, have soon been noted as being difficult for statistical word alignment: Brown et al. (1993) modelled *fertility*, as they called it, statistically in the more sophisticated IBM-models. Other approaches adopt again a normalizing procedure: in a preprocessing step, multiwords are either recognized as such and subsequently treated as if they were a single token (Tiedemann, 1999), or, reversely, the tokens they align to may be split into their components, with the components being aligned to the parts of the corresponding multiword expression on a 1:1 basis.

The latter approach is clearly insufficient for word alignment quality: it assumes that compositionality holds for both the multiword unit and its translation, i.e. that the meaning of the whole unit is made up of the meaning of its part. This clearly need not be the case, and further problems arise when a multiword unit and its translation contain different numbers of elements.

The former approach, i.e. of recognizing multiword units as such and treating them as a single token, depends on the kind of recognition procedure adopted, and on the way their alignment is

computed: if it is based on statistics, again, the approach will hardly perform well for rare expressions.

To sum up, aligning at the sentence level can be done with success using a combination of language-independent methods. Word alignment, on the other hand, still leaves room for improvement: current models do not suffice to align rare words and multiword units, and syntactic differences between source and target languages, too, still present a challenge for most word alignment strategies.

4 An alternative text alignment system

In order to address these problems, we have designed an *alternative text alignment system*, called ATLAS, that computes text alignment based on a combination of linguistically informed rules and statistical computation. It takes a linguistically annotated corpus as input[2]. The output of the text alignment system consists of the corpus alignment information and a bilingual dictionary.

During the alignment process, hypotheses on translation pairs are computed by different *alignment modules*, and assigned a *confidence value*. These hypotheses may be about paragraphs, sentences, words, or phrases.

All hypotheses are reused to refine and complete the text alignment, and in a final filtering step, implausible hypotheses are filtered out. The remaining hypotheses constitute the final overall text alignment and are used to generate a bilingual dictionary (see figure 1 for an illustration).

4.1 Core Component

The alignment process is controlled by a core component: it manages all knowledge bases, i.e.

- information contained in a system-internal dictionary,
- corpus information like the positions of tokens and their annotations, and
- the set of alignment hypotheses.

[2]The linguistic annotation currently supported includes lemmas, parts of speech, and syntactic phrases, along with information on sentence or paragraph boundaries. The annotation may include sentence alignment information, and a bilingual dictionary may be used, too.

Additionally, the core component triggers the different *alignment modules* depending on the type of a hypothesis: if, for example, a hypothesis is about a sentence pair, then the word alignment modules of ATLAS are started in order to find translation pairs within the sentence pair.

The alignment modules are run simultaneously, but independently of each other, i.e. an alignment hypothesis may be generated several times, based on cues used by different alignment modules. A word pair e.g. may be aligned based on orthographic similarity by one module, and based on syntactic information by another module.

Each hypothesis is assigned a confidence value by the alignment module that generated it, and then returned to the core component. The confidence value of each hypothesis is derived from i) its probability or similarity value, and ii) the confidence value of the parent hypothesis.

The core component may change the confidence value of a hypothesis, e.g. if it was generated multiple times by different alignment modules, based on different alignment cues. This multiple generation of the same hypothesis is taken as indication that the hypothesis is more reliable than if it had been generated by only one alignment module, and hence its confidence value is increase.

The core component adds all new information to its knowledge bases, and hands it over to appropriate alignment modules for further computation.

The process is iterated until no new hypotheses are found. Then, the core component assembles the best hypotheses to compute a final hypothesis set: starting with the hypothesis that has the highest confidence, each next-best hypothesis is tested whether it fits into the final set; if there is a contradiction between the hypotheses already in the set and the next-best, the latter is discarded from the knowledge base. If not, then it is added to the final set. This process is iterated until all hypotheses have been either added to the final hypothesis set, or have been discarded.

Cleaning-up procedures ensure that corpus items left unaligned are either aligned to null, or can be aligned based on a process of elimination: if two units a and b are contained within the same textual unit, e.g. within the same paragraph, and aligning them would not cause a contradiction with the hypotheses in the final set, then they are aligned. Finally, all remaining hypothesis are used to generate the overall text alignment, and to com-

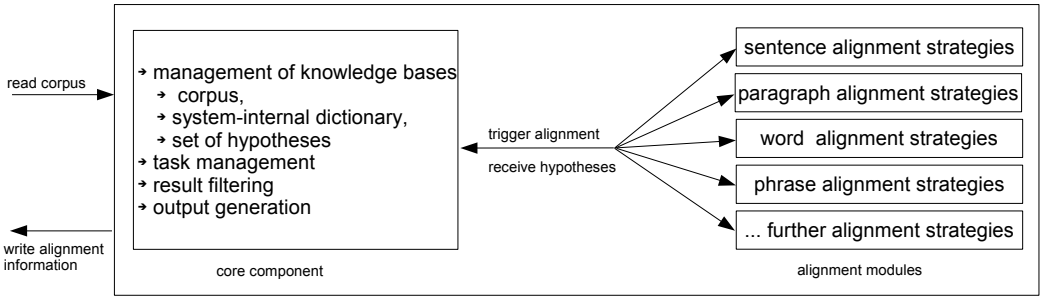

Figure 1: A schema of the text alignment architecture

pute a bilingual dictionary.

4.2 Alignment modules

Each alignment module receives a parent hypothesis as input that covers certain units of the corpus, i.e. a hypothesis on a sentence pair covers those tokens along with their annotations that are contained within the sentence pair. It uses this information to compute child hypotheses within the units of the parent hypothesis, assigns each child hypothesis a confidence value that indicates how reliable it is, and returns the set of children hypotheses to the core component.

In the case of a statistics-based alignment module, the confidence value corresponds to the probability with which a translation pair may be aligned. In other, non-statistical alignment modules, the confidence value is derived from the similarity value computed for a specific translation pair.

The alignment modules that are currently used by our the system are modules for aligning sentences or paragraphs based on the strategies that have been proposed in the literature (see overview in section 2.1), but also strategies that we have experimented with for aligning words based on linear ordering, parts of speech, dictionary lookup etc (see section 5). No statistical word alignment procedure has yet been added to the system, but we are experimenting with using statistical co-occurrence measures for deriving word correspondences. One language independent alignment strategy is based on *inheritance*: if two units a and b are aligned, then this information is used to derive alignment hypotheses for the elements within a and b as well as for the textual units that contain a and b.

5 Advantages of the hybrid architecture

As our alignment architecture is hybrid and hence need not rely on statistial information alone, it can be used to successfully address word alignment problems. Note that although linguistically informed alignment strategies are used, the system is not restricted to statistics-free computation: it is still possible to compute word co-occurrence statistics and derive alignment hypotheses.

5.1 Rare words

Linguistically-informed rules that compute alignments based on corpus annotation, *but not* on statistics, can be used to overcome data sparseness. Syntactic categories e.g. give reliable alignment cues as lexical categories such as nouns and verbs are not commonly changed during the translation process. Even if category changes occur, it is likely that the categorial class stays the same. Ideally, a noun e.g. will be translated as a noun, or if it is not, it is highly probable that it is translated as an adjective or verb, but not as a functional class member like a preposition.

Likewise, dictionary lookup may be used, and is used by or system, to align words within sentences or phrases. We have also implemented a module that aligns sentences and words based on string similarity constrained by syntactic categories: the module exploits the part of speech annotation to align sentences and words based on string similarity between nouns, adjectives, and verbs, thus modifying the classic approach by Simard et al (1992). The advantage of the modification is that the amount of cognates within lexical class words will be considerably higher than between prepositions, determiners, etc., hence filtering by word

category yields good results.

5.2 Word Order Differences

As ATLAS supports the alignment of phrases, mismatches between the linear orderings of source and target language words become irrelevant. Additionally, phrase alignment can considerably narrow down the search space within which to find the translation of a word. If e.g. a noun phrase has already been aligned to its equivalent in the other language, aligning its daughter nodes on the basis of their syntactic categories, without any further constraints or statistical information, can be sufficient.

Furthermore, if parts of the phrase can be aligned using the system-internal dictionary, aligning the remaining words could be done by process of elimination.

5.3 Multiword alignment

Multiwords are traditionally hardest to align, one reason being that they are hard to recognize statistically. With our text alignment system, however, it is possible to write i) language-specific rules that detect multiwords and define ii) a similarity measure that aligns the detected multiwords to their translations. This similarity measure may be language-pair specific, or it may be defined globally, i.e. it will be used for any language pair.

We have already tested such a procedure for aligning English nominal multiwords with their German translations: In this procedure, English nominals are detected based on their typical part-of-speech patterns, and aligned to German nouns if the two expressions are roughly of the same length, counted in characters. The results are encouraging, indicating that nominals can be aligned reliably irrespective of their frequencies in the corpus (Schrader, 2006).

6 Data

As development corpus, we are using *Europarl*, a corpus of European Parliament debates (Koehn, 2005). *Europarl* consists of roughly 30 million tokens per language and is tokenized and sentence-aligned. For the purposes of testing ATLAS, we have POS-tagged and lemmatized the German, English, and French parts of the corpus using the freely available *tree-tagger* (Schmid, 1994). Additionally, we have chunked the German and English texts with an extension of this tool (Schmid, unpublished). Table 1 shows the number of tokens and types of the corpus for all three languages. It also shows the percentages of hapax legomena, rare events[3], and all other types of the corpus.

7 Evaluation

For evaluating of our text alignment system, we are currently setting up an English-German gold standard: we have randomly chosen a debate protocol of the *Europarl* corpus that contains approximately 100,000 tokens per language (see table 2), and we corrected its sentence alignment manually. The correction was done by two annotaters independently of each other, and remaining sentence alignment differences after the corrections were resolved.

In a second step, we have chosen 242 sentence pairs from this reference set to create a word alignment gold standard. Some sentence pairs of this set have been chosen randomly, the others are taken from two text passages in the protocol. We had considered choosing sentence pairs that were distributed randomly over the reference set, however, we decided for taking complete text passages in order to make manual annotation easier. This way, the annotators can easily access the context of a sentence pair to resolve alignment ambiguities.

Additionally, we have created word alignment guidelines based on those already given by Melamed (1998) and Merkel (1999). We have annotated all 242 sentence pairs twice, and annotation differences are currently being resolved.

As this gold standard can only be used to evaluate the performance of English-German word alignment, we will also evaluate our system on the Stockholm parallel treebank (Volk and Samuelsson, 2004). Evaluating against this manually constructed treebank has the advantage that we can evaluate phrase alignment quality, and that we can gather evaluation data for the language pairs English-Swedish and Swedish-German.

We have decided to use the evaluation metrics precision, recall and the *alignment error rate* (AER) proposed by Och and Ney (2000) in order to compare results to those of other alignment systems.

[3]We define rare events here as types occurring 2 to 10 times

Language	Tokens	Types	Hapax Legomena	Rare Events	Frequent Types
English	29.077,024	101,967	39,200 (38.44%)	35,608 (34.92%)	27,159 (26.64%)
German	27.643,792	286,330	140,826 (49.18%)	98,126 (34.27%)	47,378 (16.55%)
French	32.439,353	114,891	42,114 (36.66%)	41,194 (35.84%)	31,583 (27.49%)

Table 1: Corpus characteristics of the Europarl corpus

Language	Tokens	Types	Hapax Legomena	Rare Events	Frequent Types
English	111,222	7,657	3,474 (45.37%)	3,027 (39.53%)	1,156 (15.10%)
German	91,054	11,237	6,336 (56.39%)	3,973 (35.36%)	928 (8.26%)

Table 2: Characteristics of the evaluation suite

8 Summary

Summing up, we have presented a new text alignment architecture that makes use of multiple sources of information, partly statistical, partly linguistics-based, to align bilingual, parallel texts. Its input is a linguistically annotated parallel corpus, and corpus annotation may include information on syntactic constituency, syntactic category membership, lemmas, etc. Alignment is done on various levels of granularity, i.e. the system aligns simultaneously at the paragraph, sentence, phrase, and word level. A constrained best-first search is used to filter out errors, and the output of the system is corpus alignment information along with a bilingual dictionary, generated on the basis of the text alignment.

As our system need not rely on statistics alone, the alignment of hapax legomena and other rare events is not a problem. Additionally, specific strategies have been implemented, and further can be added, to deal with various kinds of multiword units. Finally, as the system allows phrase alignment, it stands on equal footing with other phrase alignment approaches.

Currently, the system is tested on the English-German parts of the *Europarl corpus*, but as it is highly modular, it can easily be extended to new language pairs, types of information, and different alignment strategies.

First performance test have been promising, and we are setting up a gold standard alignment for a thorough evaluation.

9 Further work

We are currently adding Swedish and French to the set of supported languages, such that our system will be able to align all possible pairings with the languages German, English, French and Swedish. If possible, we want to conduct experiments that involve further languages and additional kinds of corpus annotation, like e.g. detailed morphological information as annotated e.g. within the *CroCo* project (Neumann and Hansen-Schirra, 2005).

At the same time, we are constantly extending the set of available alignment strategies, e.g. with strategies for specific syntactic categories or strategies that compute alignments based on statistical co-occurrence.

A first evaluation of our text alignment system will have been completed by autumn 2006, and we plan to make our gold standard as well as our guidelines available to the research community.

Acknowledgement

We thank Judith Degen for annotation help with the gold standard.

References

Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer. 1991. Aligning sentences in parallel corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 169–176, Berkeley, California, USA.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Colin Cherry and Dekang Lin. 2003. A probability model to improve word alignment. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 88–95, Sapporo, Japan.

Pascale Fung and Kenneth W. Church. 1994. K-vec: a new approach for aligning parallel texts. In *Proceedings of the 15th International Conference on*

Computational Linguistics (COLING), pages 1096–1102, Kyoto, Japan.

Pascale Fung and Kathleen McKeown. 1994. Aligning noisy parallel corpora across language groups: word pair feature matching by dynamic time warping. In *Proceedings of the First Conference of the Association for Machine Translation in the Americas (AMTA-94)*, pages 81–88, Columbia, Maryland, USA.

William A. Gale and Kenneth W. Church. 1991. A program for aligning sentences in bilingual corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 177–184, Berkeley, California, USA. Reprinted 1993 in Computational Linguistics.

William A. Gale and Kenneth W. Church. 1993. A program for aligning sentences in bilingual corpora. *Computational Linguistics*, 19(1):75–102.

D. Hiemstra. 1996. Using statistical methods to create a bilingual dictionary. Master's thesis, Universiteit Twente.

Martin Kay and Martin Röscheisen. 1993. Text-translation alignment. *Computational Linguistics*, 19(1):121–142.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT Summit*.

Jonas Kuhn. 2004. Exploiting parallel corpora for monolingual grammar induction – a pilot study. In *Workshop proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, pages 54–57, Lisbon, Portugal. LREC Workshop: The Amazing Utility of Parallel and Comparable Corpora.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT Press, Cambridge, Massachusetts, London.

I. Dan Melamed. 1998. Annotation style guide for the BLINKER project. Technical Report 98-06, Institute for Research in Cognitive Science, University of Pennsylvania.

Magnus Merkel. 1999. Annotation style guide for the PLUG link annotator. Technical report, Linköping university, Linköping, March. PLUG report.

Stella Neumann and Silvia Hansen-Schirra. 2005. The CroCo project. Cross-linguistic corpora for the investigateon of explicitation in translation. In *Proceedings of the Corpus Linguistics Conference*, Birmingham, UK.

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, Hong Kong, China.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*, pages 44–49, Manchester, England.

Helmut Schmid. unpublished. The IMS Chunker. unpublished manuscript.

Bettina Schrader. 2006. Non-probabilistic alignment of rare German and English nominal expressions. In *To appear in: Proceedings of the Fifth Language Resources and Evaluation Conference (LREC)*, Genoa, Italy. to appear.

Michel Simard, G. F. Foster, and P. Isabelle. 1992. Using cognates to align sentences in bilingual corpora. In *Proceedings of the Fourth International conference on theoretical and methodological issues in Machine translation*, pages 67–81, Montreal, Canada.

Frank Smadja, Kathleen R. McKeown, and Vasileios Hatzivassiloglou. 1996. Translating collocations for bilingual lexicons: A statistical approach. *Computational Linguistics*, 22(1):1–38.

Jörg Tiedemann. 1999. Word alignment - step by step. In *Proceedings of the 12th Nordic Conference on Computational Linguistics*, pages 216–227, Trondheim, Norway.

Jörg Tiedemann. 2003. Combining clues for word alignment. In *Proceedings of the 10th Conference of the European Chapter of the ACL (EACL03)*, pages 339 – 346, Budapest, Hungary.

Kristina Toutanova, H. Tolga Ilhan, and Christopher D. Manning. 2002. Extensions to HMM-based statistical word alignment models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 87–94, Philadelphia, USA.

Patrick Tschorn. 2002. Automatically aligning English-German parallel texts at sentence level using linguistic knowledge. Master's thesis, Universität Osnabrück.

Stephan Vogel, Hermann Ney, and Christoph Tillmann. 1999. HMM-based word alignment in statistical translation. In *Proceedings of the International Conference on Computational Linguistics*, pages 836–841, Copenhagen, Denmark.

Martin Volk and Yvonne Samuelsson. 2004. Bootstrapping parallel treebanks. In *Proceedings of the Workshop on Linguistically Interpreted Corpora (LINC) at COLING*, Geneva, Switzerland.