Inferring syntactic rules for word alignment through Inductive Logic Programming

Sylwia Ozdowska*, Vincent Claveau**

* CLLE-ERSS, Université de Toulouse-Le Mirail
5 allées Antonio Machado
Toulouse, France
ozdowska@univ-tlse2.fr

** IRISA - CNRS
Campus de Beaulieu
Rennes, France
Vincent.Claveau@irisa.fr

Abstract

This paper presents and evaluates an original approach to automatically align bitexts at the word level. It relies on a syntactic dependency analysis of the source and target texts and uses a machine-learning technique, namely inductive logic programming, to automatically infer rules called syntactic alignment rules. These rules make the most of the syntactic information to align words. This machine learning approach is entirely automatic, requires a very small amount of training data, and its performance rivals some of the best existing alignment systems. Moreover, syntactic isomorphisms between the source language and the target language are easily identified through the inferred rules.

1. Introduction

Sub-sentential alignment determines translational correspondences at word and phrase level given a sentence-aligned bilingual corpus. It is a fundamental component in a number of cross-language applications such as statistical machine translation (SMT) or bilingual lexicon extraction. Both the quality and quantity of word and phrase alignments have a significant effect on these tasks: the better the alignments, the better the phrase pairs for phrase-based SMT for instance. There is a range of different approaches to sub-sentential alignment. The most common approach is generative models which consist of a large number of parameters estimated in an unsupervised manner on a large bilingual corpus. The most often used implementation of generative models is Giza++ (Och and Ney, 2003)[1]. It implements the models initially developed at IBM (IBM models 1 to 5) and some of the extensions to these models (Vogel et al., 1996). In general generative models have been shown to have powerful modeling capabilities and can produce high-quality alignments. However their lack of flexibility makes it almost impossible to incorporate linguistic features. Discriminative word alignment models were developed to overcome this shortcoming (Liu et al., 2005; Moore, 2005; Ma et al., 2008). Such models can incorporate various features encoded in the input data (such as part-of-speech tags or syntactic dependency relations). On the other hand, they are trained in a supervised manner, meaning that they require (human-)annotated word alignment data for training. Another class of approaches to sub-sentential alignment operate on syntactically annotated data, either on the source side or both on source and target sides, in order to align tree structures. This includes Inversion Transduction Grammar (Wu, 2000), which performs synchronous parsing on bilingual sentence pairs to establish translational correspondences, tree-to-string alignment (Yamada and Knight, 2001; Lin and Cherry, 2003), which aligns a source tree to a target string, or tree-to-tree alignment (Ding et al., 2003; Tinsley et al., 2007; Hearne et al., 2008), which aligns a source tree to a target tree. The idea behind these syntactically-informed approaches is to restrict alignment possibilities to those that are allowed by a given syntactic representation: a bilingual synchronous grammar in (Wu, 2000), a constituency analysis in (Yamada and Knight, 2001; Lin and Cherry, 2003; Tinsley et al., 2007) and a dependency analysis in (Ding et al., 2003; Hearne et al., 2008). Syntactic information has also been used in a heuristics-based method which expands an anchor alignment using a set of manually defined syntactic alignment rules (Ozdowska, 2006)[2].

We build on this approach (Section 2.) to propose a novel alignment technique where the syntactic alignment rules are inferred in a semi-supervised manner using a symbolic machine-learning technique (namely, inductive logic programming or ILP) applied to automatically annotated word alignment data (Section 3.). Our experimental objectives are (1) to demonstrate that our symbolic approach requires far less training data to produce high quality alignments compared to the above-mentioned statistical approaches, (2) to test the generality of our model on various corpus types, and (3) to show that our approach makes it possible to interpret the alignment rules from a linguistic point of view to study the cases of isomorphisms (directly corresponding syntactic structures in the source and target languages) and non-isomorphisms (Sections 4. and 5.).

2. Syntactic alignment rules

We take as a starting point the idea of projecting alignment links within a sentence-aligned parsed bilingual corpus based on an anchor alignment and a set of syntactic alignment rules. In Figure 1 for instance, given the anchor pair *Community* / *Communauté*, it is possible to align *ban* / *a interdit* using the subject relationship.

Each rule is defined according to the syntactic relationships identified on both source and target sides of the corpus and the direction followed to project the alignment

[1] GIZA++ is available at http://www.jfoch.com/GIZA++.html.

[2] This idea has further been transposed to the above-mentioned discriminative framework (Ma et al., 2008).

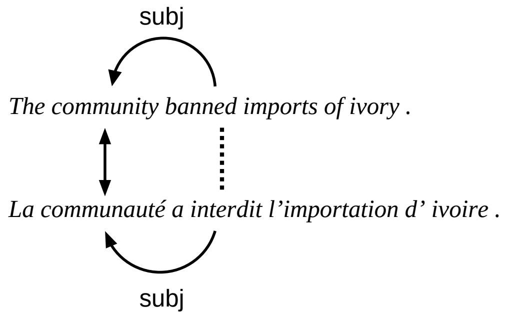

Figure 1: Syntax-based alignment

link (head-to-dependent or dependent-to-head). For instance, the rule used in Figure 1 defines a projection which goes from an anchor pair composed of nouns (*Community* / *Communauté*) and follows the subject relationship (subj) to reach their respective head (*ban* / *a interdit*):

$$\mathrm{V} \xrightarrow{\mathsf{subj}} \textbf{Noun} \;/\!\!/\; \mathrm{V} \xrightarrow{\mathsf{subj}} \textbf{Noun}$$

This rule-based approach is reported to produce a high precision alignment but recall is inferior compared with statistical approaches because the majority of the rules are defined assuming a direct correspondence between French and English (Ozdowska, 2006). Moreover, the human expertise required to define the alignment rules can be put forward as a criticism as far as extension to new language pairs or corpora is concerned. The machine learning technique we propose here has been developed with the specific intention to overcome these shortcomings.

3. Machine Learning

The novelty of our approach lies in using symbolic machine learning to automatically infer syntactical alignment rules. The principle behind this approach is the following: from examples of valid word alignments in a sentence pair, we automatically learn rules that represent the syntactic structures these alignments are embedded into. To do this, we use Inductive Logic Programming (ILP). In the following sub-sections, we present ILP and describe how to apply it to infer syntactic alignment rules.

3.1. Inductive Logic Programming

ILP is at the intersection of machine learning and logic programming. It makes it possible to infer a set H of general rules describing a concept from a set E^+ of examples of this concept (and possibly a set E^- of counter-examples), and external information called *Background Knowledge* and noted B (an in-depth presentation of ILP can be found in (Muggleton and De Raedt, 1994)). The examples and the background knowledge are described in a logical language (predicate logic such as Prolog); the inferred rules are Horn Clauses, obtained by generalizing the examples using the background knowledge information.

Several conditions constraint the learning process; they form the logical framework of ILP. The two following conditions apply to the data ($\Box$ means *false* and $\models$ represents the logical implication):

- *a priori* correctness ensures that counter-examples cannot be deduced from the *Background Knowledge*, that is $B \wedge E^- \not\models \Box$;
- *a priori* necessity emphasizes the need of additional knowledge –the rules to infer– to explain the examples, that is $B \not\models E^+$.

Two additional conditions concern the set of rules H we want to infer:

– *a posteriori* correctness requires that counter-examples cannot be deduced from B and H, that is: $B \wedge H \wedge E^- \not\models \Box$;

– completeness ensures that every positive example is explained by the rules in H: $B \wedge H \models E^+$.

In practice, the rules in H are searched for in a huge (sometimes infinite) hypothesis space containing all possible rules. This space is usually hierarchicaly structured, which makes an efficient search possible. A rule from this space is kept in H if it maximizes a score usually defined with respect to the number of examples (and possibly of counter-examples) this rule covers.

Because of its expressiveness, ILP has already been used for various machine learning tasks, especially in NLP (Cussens et al., 2000; Claveau et al., 2003, *inter alia*). We have called it upon for alignment because syntactic dependency relations and translation relationships can intuitively be encoded with predicates (see below). Moreover, compared to existing statistical or machine learning approaches, ILP has the advantage of inferring rules which are easy to interpret and which allow a linguistic analysis of isomorphisms and non-isomorphisms across languages.

3.2. Inferring Alignment Rules

The rules we want to infer are syntactic alignment rules. The examples we use are parsed aligned sentences containing valid anchor alignments. In B, we encode syntactic dependency relations and anchor alignments used as examples. For instance, if we know that *companies* // *entreprises* can be aligned in the following sentence pair (the identifier of each word is noted after the slashes):

... *private/id_1_en* ***sector**/id_2_en* *companies/id_3_en*
... *les/id_1_fr entreprises/id_2_fr du/id_3_fr* ***secteur**/id_4_fr* *privéid_5_fr*

E^+ is populated with the predicate:
alignment(id_3_en,id_2_fr).
and B is populated with the predicates (the name of the predicate represents the name of the syntactic relation; here, the two arguments are constants: the head's and the dependent's identifiers):

```
det(id_2_fr,id_1_fr).          prep(id_2_fr,id_3_fr).
cprep(id_3_fr,id_4_fr).        adj(id_4_fr,id_5_fr).
adj(id_2_en,id_1_en).          nn(id_3_en,id_2_en).
anchor(id_2_en,id_4_fr).
```

One example of a valid rule that may be inferred from this example is (this time, the arguments are variables that can represent any word identifier):

alignment(M_en,M_fr) :- nn(M_en,A1), prep(M_fr,F1), cprep(F1,F2), anchor(A1,F2).

This rule descibes the correspondence between Noun Noun structures in English and Noun de Noun structures in French and is equivalent to:

$$\text{M\_en} \xrightarrow{\text{nn}} \mathbf{A1} \mathbin{/\!/} \text{M\_fr} \xrightarrow{\text{prep}} \text{F1} \xrightarrow{\text{cprep}} \mathbf{F2}.$$

Every time such a configuration is found in a sentence pair, the words that appear in it can be aligned.

Since no counter examples are used, it is important to ensure that our learning technique do not over-generalize the examples, meaning that the inferred rules only cover correct alignments. To do so, we impose that a rule has to cover a minimum number of examples from E^+ in order to be kept in H; this number is called the *coverage threshold*. In addition, another mechanism called language biases is used to reject rules that might be ill-formed. These biases are part of every Machine Learning technique (Mitchell, 1980); within the ILP framework, they ensure that only well-formed rules are inferred. In our case, a rule is well-formed if it describes a path linking an anchor alignment to two unaligned words through a syntactic relationship. Only the rules respecting these two criteria are finally proposed by our ILP algorithm and used hereafter to perform alignment.

4. Data and tools

The experiments we present are based on the French-English data containing the Canadian Hansards that was provided within the framework of the HLT'03 word alignment evaluation campaign (Mihalcea and Pedersen, 2003). We used 10 to 1 000 sentence pairs for training out of 1.3 million sentence pairs made available. The test set contained 447 human-annotated sentences with alignments marked S (alignment considered as sure across annotators) or P (otherwise). Figure 2 gives an example of a reference sentence pair (plain links represent S alignments; dashed links represent P alignments).

On the other hand, in order to test the genericity of the rules with respect to the type of training corpus, we also inferred alignment rules from two additional French-English corpora: INRA and JOC. INRA is a corpus of research and popular science articles[3] whereas JOC contains texts issued by the European Commission[4]. We used 1 000 sentence pairs of each corpus for training and the same test set as described above for evaluation.

All the data was dependency-parsed using Syntex (Fabre and Bourigault, 2001) for both French and English. Anchor words were obtained automatically by a simple bootstrapping technique based on similarity functions and cognates. On average 4 to 6 anchor pairs were detected by sentence pair with a minimal precision of 85%, depending on the corpus. Each anchor pair was used as a positive alignment example in the sentence pair where it appeared to learn syntactic alignment rules through ILP. This bootstrapping technique makes our approach entirely automatic. The rules inferred could then be applied to new bi-texts containing anchor alignments.

5. Results

5.1. Evaluation methodology

Our approach is evaluated against the HLT test set. In the first experiment, alignment rules are inferred from the HANSARD corpus. We compute recall and precision rates along with f-measure to measure performance.

We also show the results of the best-performaing alignment systems (in terms of f-measure) that took part in the HLT campaign to allow comparison: Ralign (Simard and Langlais, 2003), XRCE (Déjean et al., 2003), BiBr (Zhao and Vogel, 2003) and ProAlign (Lin and Cherry, 2003) systems. In addition to this, we give the results of the ALIBI system (Ozdowska, 2006) in which the alignment rules were manually defined.

5.2. Alignment performance

Table 1 presents the results for the S alignment task. For this experiment, the ILP-based learning step was carried out on a training corpus composed of 1 000 bi-sentences from the HANSARD (not contained in the test set).

Our ILP-based approach achieves reasonable results in comparison with other approaches. In particular, it yields higher precision because syntactic annotation on both source and target sides reinforces the cohesion constraint and limits alignment to syntactically motivated links. A detailed analysis of the results shows that most errors are caused either because of missing anchor pairs or missing dependency relations. The latter are particularily harmful for recall since alignment links cannot be projected to words that are not syntactically connected within a sentence.

Despite the fact that our ILP-based system (as well as ALIBI) is more specifically designed to perform non ambiguous alignments (S alignments), we also present results obtained for P alignments in Table 2. Like in the previous table, we also indicate the other systems' performance on this task.

Several things are noteworthy in this table. First, the ILP system's results are better in terms of f-measure than those of its manual counterpart ALIBI, which suffers from a poor recall rate. It means that our machine learning process yields rules that cover more syntactic structures but are slightly less precise than the manually defined rules operating in ALIBI. However, the recall achieved with ILP is too low to be competitive with regard to the other systems. This is not surprising since syntactical constraints do not allow to produce ambiguous links such as an alignment between a determiner and a verb for example. The ProAlign system, which also uses syntactic constraints, suffers from the same problem when compared with the three other systems based on a purely statistical approach. Moreover, decomposing mappings involving multi-word expressions on the source and/or target side into one-to-one P alignments is an issue as it might not be the best way to reflect systems' performance, in particular as far as recall is considered (Ozdowska, 2006). Such a decomposition results in an over-generation of reference alignments compared to what

[3] Provided by A. Lacombe from INRA.

[4] Provided by the ARCADE project (Véronis, 2000).

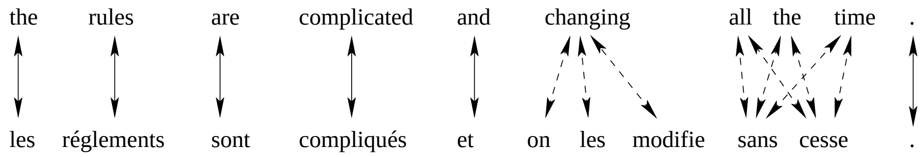

Figure 2: Annotation for the HLT alignment campaign

System	ILP	ALIBI	Ralign	XRCE	BiBr	ProAlign
Precision	82.11%	**88.78%**	72.54%	55.54%	63.03%	71.94%
Recall	74.09%	66.86%	80.61%	**93.46%**	74.59%	91.48%
F-measure	77.89%	76.28%	76.36%	69.68%	68.32%	**80.54%**

Table 1: Performance of the ILP-based alignment system on the HLT dataset on S alignments

systems are actually able to produce. For example, a translation relation such as *all the time // sans cesse* is represented with 6 P alignments.

5.3. Size of the training corpus

In this section, we are interested in the evolution of performance with respect to the size of the training corpus. We carry out several experiments in which we only change the amount of training sentence pairs (extracted from the HANSARD corpus) used to infer alignment rules. Figure 3 shows the recall rate, precision rate and f-measure obtained according to the amount of training data. Almost no variation in recall and precision rates is observed for a training corpus ranging from 300 to 1 000 sentence pairs. Below 300 sentence pairs, the precision rate is slightly higher while the recall rate is lower. A possible explanation is that only few alignment rules, the most reliable rules, are inferred. Finally, it is worth noting that 10 sentence pairs are sufficient to infer rules that are relevant enough to yield 70% f-measure. These results can be considered as very positive with respect to the very small size of the training corpus. As a comparison, probabilistic systems like Ralign, XRCE, BiBr and ProAlign use 1.3 million training sentence pairs.

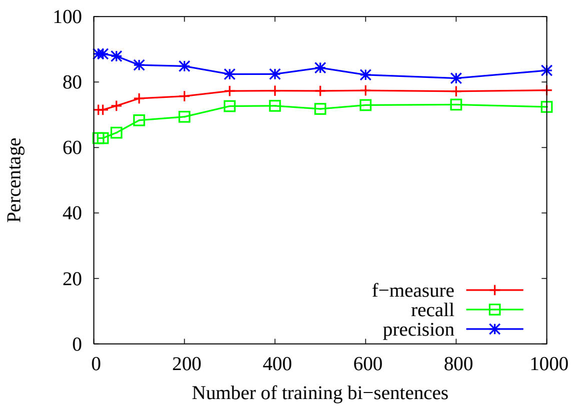

Figure 3: Performance variations with respect to the number of training bi-sentences

5.4. Examination of the results

Most alignment errors our systems make can be classified into a limited number of categories. As we previously said, the lack of anchor pairs and of dependency relations accounts for most false negatives (*i.e.* non-detected alignments).

Some of the false positives (misalignments) simply result from parsing errors (which can be caused by part-of-speech tagging mistakes). For example, consider the sentence pair: *federal government carpenters get $ 6.42 // Les menuisiers du gouvernement fédéral touchent $ 6.42*. The adjective *federal* was wrongly attached by Syntex to *carpenters* leading to the misalignment *carpenter // gouvernement*, both words being tagged as heads of the anchor pair *federal // fédéral*.

On the other hand, certain rules bring up noisy alignments because they are not specific enough. For example, information pertaining to voice is not encoded in the rules so that it is impossible to distinguish cases where a subject is to be aligned to the object because of a active/passive voice shift in the translation. This is why *gouvernement* and *legislation* are misaligned in the sentence pair (the anchor alignment is *bring // apporter*): *good legislation has been brought in by Liberal governments // les gouvernements libéraux ont apporté de bonnes mesures législatives*.

Finally, rephrasings interfere with the alignment process. For instance, in the sentence pair *the Government must implement the recommendations of the Commissioner of Official Languages // le gouvernement se doit de respecter les recommandations du Commissaire aux langues officielles*, *implement* has been aligned to *respecter* while this alignment is not present in the HLT testset.

5.5. Inferred rules

With 1 000 training sentences, 81 rules are inferred from the INRA corpus, 82 from JOC and 65 from HANSARD with a cover threshold at least equal to 2 (*cf.* Section 3.2.). These numbers respectively fall to 67, 62 and 54 if the coverage threshold is set to 10, which was the case in our experiments. For both thresholds, about 40 rules are common to all 3 corpora. 21 rules with a coverage threshold higher than 10 are specific to the INRA corpus, 13 to JOC, and 13 to HANSARD. Considering all inferred rules, the majority

System	ILP	ALIBI	Ralign	XRCE	BiBr	ProAlign
Precision	80.65%	90.81%	77.56%	89.65%	66.11%	**96.49%**
Recall	28.44%	22.49%	**38.19%**	34.92%	30.06%	28.41%
F-measure	42.05%	36.05%	**51.18%**	50.27%	41.33%	43.89%

Table 2: Performance of the ILP-based alignment system on the HLT dataset on P alignments

is similar to those manually defined in ALIBI (Ozdowska, 2006).

Most of the rules describe common syntactic isomorphisms across English and French, such as the alignment of adjectives modifying two previously aligned nouns or the alignment of direct objects of two previously aligned verbs:

alignment(M_en,M_fr) :- adj(C,M_en), adj(D,M_fr), anchor(C,D).

alignment(M_en,M_fr) :- obj(C,M_en), obj(D,M_fr), anchor(C,D).

These cases of perfect isomorphisms represent almost 50% of the inferred alignment rules. Some rules accounting for non-isomorphic structures are also found, like for instance the standard configuration of nominal phrases: Noun Noun in English and Noun de Noun in French (*cf.* Section 3.2.). Certain non-isomorphic rules can even lead to aligning of words with different parts-of-speech, such as nouns and adjectives:

alignment(M_en,M_fr) :- nn(C,M_en), adj(D,M_fr), anchor(C,D).

So far, the analysis of the rules makes it clear that they are mostly generic alignment patterns. Their similarity with the rules manually defined in (Ozdowska, 2006) is a strong indication that our machine learning approach is well-grounded. Nonetheless, some rules are unexpected — and their validity can be questionned— like for instance:

alignment(M_en,M_fr) :- adj(M_fr,C), nn(D,M_en), adj(D,E), anchor(E,C).

This rule allows the alignment of *bargaining* and *négociation* in the sentence pair *[...] to have some hang-up with regard to the* ***collective*** *bargaining process // [...] éprouver certains complexes à l'égard de la négociation* ***collective***.

5.6. Genericity of the syntactic alignment rules

As it was previously indicated, it is interesting to study whether or not the rules we infer are specific to a training corpus. To get an idea of their genericity from a quantitative point of vue, we first compare the alignments obtained with the set of rules inferred from each corpus against the same test set (HLT test set). Here again, training was carried out on 1 000 sentence pairs extracted from each corpus. Table 3 presents the results yield with each set of rules.

We note that the 3 sets of rules perform almost equally on the HLT test set. The precision rates are very close, with a slight advantage for the INRA rules. Conversely, this set of rules has the lowest recall rate; the best rate is achieved with the rules inferred from the JOC corpus. Unsurprisingly, the best perfomance in terms of f-measure is obtained with the rules inferred from the same corpus as the test set (yet a different part of it), that is HANSARD.

From a more qualitative point of view, we can see that the

Training corpus	HANSARD	JOC	INRA
Precision	82.08%	80.65%	**83.16%**
Recall	74.09%	**74.10%**	66.90%
F-measure	**77.88%**	77.20%	74.15%

Table 3: Performance of the ILP system according to the training corpus

number of rules for each corpus is almost equal (that is about 60 rules for 1 000 training sentence pairs and for a coverage threshold of 10 or higher). There are very few differences across the 3 sets of rules: the proportion of identical rules represents between 2/3 and 3/4 of all rules inferred from each corpus. As a consequence, few rules are corpus specific of which some are not valid due to parsing errors in the training data. The similarity across the 3 sets of rules clearly explains the proximity of the performance.

6. Conclusion and future work

In this paper we proposed an innovative technique using syntax and semi-supervised machine learning as key-components for word alignment. It automatically learns syntactic alignment rules from examples of aligned word pairs which are obtained through a bootstrapping method. Therefore, the entire process is fully automated. The alignment performance obtained equals some of the best performing statistical systems. Moreover, an important advantage to our approach is it yields reasonable results when trained on small amounts of sentence pairs as opposed to statistical systems: it can achieve 70% f-measure with as little training data as 10 sentence pairs. Whereas the amount of data is crucial for statistical approaches to accurately estimate alignment probabilities, our approach uses linguistic information at its core.

On the other hand, the inferred alignment rules give insights on the cases of isomorphisms and non-isomorphisms across the two considered languages. Finally, our experiments show that the syntactic alignment rules for French/English are quite generic as they almost do not vary across training corpora.

Several extensions are foreseen for this work. Since our approach is fully automated, we could test it on various syntactic annotations generated by different parsers in order to measure to what extent our approach depends on a particular syntactic representation. Concerning the machine learning technique, we would like to integrate part-of-speech information into the process. It may help achieve better precision preventing misalignments whose causes were detailed above. The quality of the inferred rules could also be improved using negative examples in the learning step. These negative examples must be word pairs which are known not to be translation of each other. They may help avoid excessive generalization and discard rules that are not precise enough to yield accurate alignments. However, the negative examples have to be generated in an unsupervised way, like the positive examples, so that our approach is kept entirely automated.

From an applicative point of view, our alignment technique could be applied to different language pairs. This requires that the two considered languages are syntactically related or at least that the equivalent structures, even if not similar, are regular enough to be learnt by ILP. Another requirement is obviously the availability of syntactic parsers for both languages. The examination of the inferred alignment rules would provide a way to study isomorphic and non-isomorphic structures cross-linguistically.

7. References

Vincent Claveau, Pascale Sébillot, Cécile Fabre, and Pierrette Bouillon. 2003. Learning Semantic Lexicons from a Part-of-Speech and Semantically Tagged Corpus Using Inductive Logic Programming. *Journal of Machine Learning Research, Special Issue on Inductive Logic Programming*, 4:493–525.

James Cussens, Sašo Džeroski, and (eds). 2000. *Learning Language in Logic*, volume 1925 of *Lecture Notes in Artificial Intelligence*. Sringer Verlag, June.

Hervé Déjean, Eric Gaussier, Cyril Goutte, and Kenji Yamada. 2003. Reducing parameter space for word alignment. In Mihalcea and Pedersen (Mihalcea and Pedersen, 2003), pages 23–26.

Yuan Ding, Daniel Gildea, and Martha Palmer. 2003. An algorithm for word-level alignment of parallel dependency trees. In *Proceedings of the 9th Machine Translation Summit of the International Association of Machine Translation*, New Orleans, LA, USA.

Cécile Fabre and Didier Bourigault. 2001. Linguistic clues for corpus-based acquisition of lexical dependencies. In *Proceedings of the Corpus Linguistics*, Lancaster University.

Mary Hearne, Sylwia Ozdowska, and John Tinsley. 2008. Comparing constituency and dependency representations for smt phrase-extraction. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN'08)*, Avignon, France.

Dekang Lin and Colin Cherry. 2003. Proalign: Shared task system description. In Mihalcea and Pedersen (Mihalcea and Pedersen, 2003).

Yang Liu, Qun Liu, and Shouxun Lin. 2005. Log-linear models for word alignment. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 459–466, Ann Arbor, MI, USA.

Yanjun Ma, Sylwia Ozdowska, Yanli Sun, and Andy Way. 2008. Improving word alignment using syntactic dependencies. In *Proceedings of the Second Workshop on Syntax and Structure in Statistical Translation (SSST-2)*, pages 69–77, Columbus, OH, USA.

Rada Mihalcea and Ted Pedersen, editors. 2003. *Proceedings of the HLT-NAACL 2003 Workshop Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, Edmonton, Alberta, Canada.

Tom M. Mitchell. 1980. The Need for Biases in Learning Generalizations. In *Readings in Machine Learning*. Morgan Kaufmann. Published in 1991.

Robert C. Moore. 2005. A discriminative framework for bilingual word alignment. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 81–88, Vancouver, Canada.

Stephen Muggleton and Luc De Raedt. 1994. Inductive Logic Programming: Theory and Methods. *Journal of Logic Programming*, 19-20.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statisical alignment models. *Computational Linguistics*, 1(29).

Sylwia Ozdowska. 2006. ALIBI*, un système d'ALIgnement BIlingue à base de règles de propagation syntaxique*. Ph.D. thesis, Université de Toulouse 2, Toulouse, France, December.

Michel Simard and Philippe Langlais. 2003. Statistical translation alignment with compositionality constraints. In Mihalcea and Pedersen (Mihalcea and Pedersen, 2003), pages 19–22.

John Tinsley, Ventsislav Zhechev, Mary Hearne, and Andy Way. 2007. Robust Language Pair-Independent Sub-Tree Alignment. In *Proceedings of MT Summit XI*, Copenhagen, Denmark.

Jean Véronis. 2000. *Parallel Text Processing*. Kluwer Academic Publishers, Dordrecht.

Stefan 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.

Dekai Wu, 2000. *Bracketing and Aligning Words and Constituents in Parallel Text using Stochastic Inversion Transduction Grammars*, chapter 7. In Véronis (Véronis, 2000).

Kenji Yamada and Kevin Knight. 2001. A Syntax-based Statistical Translation Model. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA, USA.

Bing Zhao and Stephan Vogel. 2003. Word alignment based on bilingual bracketing. In Mihalcea and Pedersen (Mihalcea and Pedersen, 2003), pages 15–18.