LAMP-TR-082 February 2002
CS-TR-4333
UMIACS-TR-2002-15

Improved Word-Level Alignment: Injecting Knowledge about MT Divergences

Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, Nizar Habash

Language and Media Processing Labratory
Instititue for Advanced Computer Studies
College Park, MD 20742

Abstract

Word-level alignments of bilingual text (bitexts) are not an integral part of statistical machine translation models, but also useful for lexical acquisition, treebank construction, and part-of-speech tagging. The frequent occurrence of divergences, structural differences between languages, presents a great challenge to the alignment task. We resolve some of the most prevalent divergence cases by using syntactic parse information to transform the sentence structure of one language to bear a closer resemblance to that of the other language. In this paper, we show that common divergence types can be found in multiple language pairs (in particular, we focus on English-Spanish and English-Arabic) and systematically identified. We describe our techniques for modifying English parse trees to form resulting sentences that share more similarity with the sentences in the other languages; finally, we present an empirical analysis comparing the complexities of performing word-level alignments with an without divergence handling. Our results suggest that divergence-handling can improve word-level alignment.

***The support of the LAMP Technical Report Series and the partial support of this research by the National Science Foundation under grant EIA0130422 and the Department of Defense under contract MDA9049-C6-1250 is gratefully acknowledged.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
FEB 2002		**00-02-2002 to 00-02-2002**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Improved Word-Level Alignment: Injecting Knowledge about MT Divergences	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Language and Media Processing Laboratory,Institute for Advanced Computer Studies,University of Maryland,College Park,MD,20742-3275**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**10**	

Improved Word-Level Alignment: Injecting Knowledge about MT Divergences

Bonnie J. Dorr, Lisa Pearl, Rebecca Hwa, Nizar Habash

February 14, 2002

Paper ID: ACL-2002-P0364

Keywords: lexical semantics, translation, multilinguality, alignment, divergences

Contact Author: bonnie@cs.umd.edu

Under consideration for other conferences (specify)? none

Abstract

Word-level alignments of bilingual text (bitexts) are not only an integral part of statistical machine translation models, but also useful for lexical acquisition, treebank construction, and part-of-speech tagging. The frequent occurrence of *divergences*, structural differences between languages, presents a great challenge to the alignment task. We resolve some of the most prevalent divergence cases by using syntactic parse information to transform the sentence structure of one language to bear a closer resemblance to that of the other language. In this paper, we show that common divergence types can be found in multiple language pairs (in particular, we focus on English-Spanish and English-Arabic) and systematically identified. We describe our techniques for modifying English parse trees to form resulting sentences that share more similarity with the sentences in the other languages; finally, we present an empirical analysis comparing the complexities of performing word-level alignments with and without divergence handling. Our results suggest that divergence-handling can improve word-level alignment.

Improved Word-Level Alignment: Injecting Knowledge about MT Divergences

Paper ID: ACL-2002-P0364

Abstract

Word-level alignments of bilingual text (bitexts) are not only an integral part of statistical machine translation models, but also useful for lexical acquisition, treebank construction, and part-of-speech tagging. The frequent occurrence of *divergences*, structural differences between languages, presents a great challenge to the alignment task. We resolve some of the most prevalent divergence cases by using syntactic parse information to transform the sentence structure of one language to bear a closer resemblance to that of the other language. In this paper, we show that common divergence types can be found in multiple language pairs (in particular, we focus on English-Spanish and English-Arabic) and systematically identified. We describe our techniques for modifying English parse trees to form resulting sentences that share more similarity with the sentences in the other languages; finally, we present an empirical analysis comparing the complexities of performing word-level alignments with and without divergence handling. Our results suggest that divergence-handling can improve word-level alignment.

1 Introduction

Word-level alignments of bilingual text (bitexts) are not only an integral part of statistical machine translation models, but also useful for lexical acquisition, treebank construction, and part-of-speech tagging (Yarowsky and Ngai, 2001). The frequent occurrence of "divergences", structural differences between languages, presents a great challenge to the alignment task. In this paper, we show that common divergence types can be found in multiple language pairs and systematically identified. We focus on English-Spanish and English-Arabic, presenting techniques for modifying English parse trees to form resulting sentences that share more similarity with the sentences in the other languages. We resolve some of the most prevalent divergence cases by using syntactic parse information to transform the sentence structure of one language to bear a closer resemblance to that of the other language.

The following three ideas motivate the development of automatic "divergence correction" techniques:

1. Every language pair has translation divergences that are easy to recognize.

2. Knowing what they are and how to accommodate them provides the basis for refined word-level alignment.

3. Improved word-level alignment results in improved projection of structural information from English to the foreign language.

This paper elaborates primarily on points 1 and 2 above, but our ultimate goal is to set these in the context of 3, i.e., for training foreign-language parsers to be used statistical machine translation.

A divergence occurs when the underlying concepts or gist of a sentence is distributed over different words for different languages. For example, the notion of floating across a river is expressed as *float across a river* in English and *cross a river floating* (*atravesó el río flotando*) in Spanish (Dorr, 1993) or similarly (عبر النهر طوفاً) in Arabic.

While seemingly transparent for human readers, this throws statistical aligners for a serious loop. Far from being a rare occurrence, our preliminary investigations revealed that divergences occurred in approximately 1 out of every 3 sentences in a sample size of 19K sentences from the TREC El Norte Newspaper Corpus[1] (using automatic detection techniques followed by human confirmation). Thus, finding a way to deal effectively with these divergences and repair them would be a massive advance for bilingual alignment.

The current avenue of research involves transforming English into a pseudo-English form (which we call E') that more closely matches the physical form of the foreign language, e.g., "float across a river" becomes "cross a river floating" if the foreign language is Spanish. Ideally, this rewriting of the English sentence creates more one-to-one correspondences which, in turn, facilitates our statistical alignment process. The key is to identify possible rewritings for known examples (our training set), which then generalize to rewritings for examples not yet covered in our training set. In theory, our rewriting approach applies to all divergence types. Thus, given a corpus, divergences are identified, rewritten, and then run through the statistical aligner of choice.

The alignment process is enhanced by the injection of linguistic knowledge into the standard parameter tables used in statistical language modeling: the *distortion* table (d)—for reordering the words of the English sentence, the *translation* table (t)—for translating the words of the English sentence, and the insertion table (n)—for inserting words into the English sentence. Although statistical alignment relies on the values of the parameters in these three tables, the process need not have any deeper knowledge other than these values. For example, the head-swapping divergence above based on the notion of "float across" would involve a reordering of the English words, as dictated by the values in the d-table, prior to alignment with the Spanish sentence containing "atravesó flotando".

The next section sets this work in the context of related work on alignment and projection of structural information between languages. Section 3 describes the range of divergence types covered in this work (with examples in Spanish and Arabic). Section 5 describes an experiment we undertook to examine the benefits of injecting linguistic knowledge into the alignment process. We present an empirical analysis comparing the complexities of performing word-level alignments with and without divergence handling. We conclude that annotators agree with each other more consistently when performing word-level alignments on bitext with divergence handling.

2 Related Work

Recently, researchers have extended traditional statistical machine translation (MT) models (Brown et al., 1990; Brown et al., 1993) to include the syntactic structures of the languages (Alshawi et al., 2000; Alshawi and Douglas, 2000; Wu, 1997). Furthermore, Yamada and Knight (2001) have shown that MT models are significantly improved when trained on syntactically *annotated* data (Yamada and Knight, 2001). The cost of human labor in producing annotated treebanks is often prohibitive, thus making the construction of such data for new languages infeasible. Some researchers have developed techniques for fast acquisition of hand-annotated Treebanks (Fellbaum et al., 2001). Others have appealed to machine learning techniques. For example, the work of Hermjakob and Mooney (1997) and Hwa (2000) aimed to minimize the amount of annotated data needed to induce a parser.

In a cross-language setting, some researchers have taken approaches that circumvent the construction of annotated treebanks, producing MT systems without parsers and dictionaries. One example is the Expedition effort, an enterprise to machine translation capability for low-density languages in short periods of time (Amtrup et al., 1999). Others have proposed to induce a parser from a noisy treebank of

[1] This analysis was done on the TREC Spanish Data, LDC catalog no LDC2000T51, ISBN 1-58563-177-9, 2000.

foreign-language dependency trees that are automatically projected from English (Hwa et al., 2002). Our approach is the most relevant to the latter in that it provides a significantly noise-reduced foreign-language dependency treebank for inducing a foreign language parser. We repair misalignments resulting from cross-language divergences, bringing about a more accurate dependency-tree projection from English into the foreign language.

3 Background: Divergences

As stated at the outset, every language pair has divergences that are easy to recognize. Our empirical work on English-Spanish and English-Arabic translation has revealed that there six divergences of interest. Each one is described, in turn, below.

3.1 Light Verb Construction

A *light verb construction* involves a single verb in one language being translated using a combination of a semantically "light" verb, i.e., it carries little or no specific meaning in its own right, and some other meaning unit (perhaps a noun) to convey the appropriate meaning. English light verbs include *give*, *make*, *do*, *take*, and *have*. Our findings indicate that Spanish tends to be more verbose (employing the Light Verb Construction) than English; by contrast, Arabic tends to be more contracted, mapping to a Light Verb Construction on the English side:

(1a) **English-Spanish:**
to kick ⇔ dar una patada (give a kick)
to end ⇔ poner fin (put end)
to note ⇔ tomar nota (take note)

(1b) **English-Arabic:**
to do well ⇔ احسن (do-good)
to be not ⇔ ليس (be-not)
to make do ⇔ اكتفى (make-do)

3.2 Manner Conflation

This divergence is involves translating of a single manner verb (e.g., *float*) as a light verb of motion and a manner-indicating content word. In Spanish, typically the content word is a progressive manner verb whereas Arabic generally involves the translation a verb into an English verb of motion and an adverbial describing an aspect of the motion (such as direction).

(2a) **English-Spanish:**
to float ⇔ ir flotando (go (via) floating)
to pass ⇔ ir pasando (go passing)

(2b) **English-Arabic:**
to take out ⇔ اخرج (take-out)
to come again ⇔ رجع (return)
to go west ⇔ غرّب (go-west)

3.3 Head Swapping

This divergence involves the demotion of the head verb and the promotion of one of its modifiers to head position. In other words, a permutation of semantically equivalent words is necessary to go from one language to the other. In Spanish, this divergence is typical in the translation of an English motion verb and a preposition as a directed motion verb and a progressive verb. This divergence is less common in the case of English-Arabic. Examples are given below.

(3a) **English-Spanish:**
to run in ⇔ entrar corriendo (enter running)
to fly about ⇔ andar volando (go-about flying)

(3b) **English-Arabic:**
to laugh the night away ⇔ امضى الليلة ضحكاً (pass-away the-night laughing)
to do something quickly ⇔ اسرع في فعل الشيء (go-quickly in doing something)

3.4 Thematic Divergence

A thematic divergence occurs when the verb's arguments switch thematic roles from one language to another. The Spanish verbs *gustar* and *doler* are examples of this case. This type of divergence is very common in Spanish and is, in fact, the most abundant divergence type in the TREC El Norte Corpus. Although thematic divergences arise in the English-Arabic case as well, it is less common. Consider the cases below.

(4a) **English-Spanish:**
I like grapes ⇔ Me gustan uvas (to-me please grapes)
I have a headache ⇔ me duele la cabeza (to-me hurt the head)

(4b) **English-Arabic:**
I like grapes ⇔ العنب يعجبني (grapes please-me)
I have a headache ⇔ رأسي يؤلمني (my-head hurt-me)

3.5 Categorial Divergence

A categorial divergence involves a translation that uses different parts of speech. In the English-Spanish example below, the adjectival phrase is translated into a light verb accompanied by a nominal version of the adjective. A common form of this divergence between English and Arabic is the nominalization of the English verb. The examples below illustrate this divergence.

(5a) **English-Spanish:**
to be jealous ⇔ tener celos (to have jealousy)
to be fully aware ⇔ tener plena conciencia (have full awareness)

(5b) **English-Arabic:**
when he returns ⇔ عند رجوعه (upon return-his)

3.6 Structural Divergence

A structural divergence involves the realization of incorporated arguments such as subject and object as obliques (i.e. headed by a preposition in a PP). The following are examples in English-Spanish and English-Arabic:

(6a) **English-Spanish:**
to enter the house ⇔ entrar en la casa (enter in the house)
ask for a referendum ⇔ pedir un referendum (ask-for a referendum)

(6b) to seek ⇔ بحث عن (search for)
God commanded it ⇔ اوصى بها الله (commanded with-it God)

4 Occurrence of Divergences in Large Corpora

In theory, the divergences illustrated in the previous section are common to every language. However, they may be realized in different ways in different language pairs. As we have seen, Spanish may be analyzed as a rewriting of English that involves "expansion," i.e., the Spanish appears to be more verbose. On the other hand, the same divergence types showed up differently in Arabic; the rewritings appear to be a "contraction" of the English, rather than "expansion," i.e., the English appears to be more verbose.

We investigated divergences in Arabic and Spanish corpora to determine how often such cases arise.[2] First, we developed a set of handcrafted regular expressions for detecting divergent sentences in Arabic and Spanish corpora. The Arabic regular expressions were derived by examining a small set of sentences (50), a process which took approximately 20 person-hours. The Spanish expressions were derived by a different process—involving a more general analysis of the behavior of the language—taking approximately 2 person-months.

We applied the Spanish and Arabic regular expressions to a sample size of 19K TREC sentences and 1K sentences from the Arabic Bible. Each automatically detected divergence was subsequently human verified and categorized into a particular divergence category. Table 1 indicates the percentage of cases we detected automatically and also the percentage of cases that were confirmed (by humans) to be actual cases of divergence.

It is important to note that these numbers reflect the techniques used to calculate them. Because the Spanish regular expressions were derived through a more general analysis of the language, the precision is higher in Spanish than it is in Arabic. Human inspection confirmed approximately 1995 Spanish sentences out of the 2109 that were automatically detected (95% accuracy), whereas whereas 124 sentences were confirmed in the 319 detected Arabic divergences (39% accuracy).

On the other hand, the more constrained Spanish expressions appear to give rise to a lower recall. In fact, an independent study with more relaxed regular expressions on the same 19K Spanish sentences resulted in the automatic detection of divergences in 18K sentences (95% of the corpus), 6.8K of which were confirmed by

[2] For Spanish, we used TREC Spanish Data; for Arabic, we used an electronic, version of the Bible written in Modern Standard Arabic—28K verses.

Language	Detected Divergences	Human Confirmed	Sample Size (sentences)	Corpus Size (sentences)
Spanish	11.1%	10.5%	19K	150K
Arabic	31.9%	12.4%	1K	28K

Table 1: Divergence Statistics

humans to be correct (35% of the corpus). Future work will involve repeated constraint adjustments on the regular expressions to determine the best balance between precision and recall for divergence detection; we believe the Arabic expressions fall somewhere in between the two sets of Spanish expressions (which are conjectured to be at the two extremes of constraint relaxation—very tight in the case above and very loose in our independent study).

5 Experiment: Impact of Divergence Correction on Alignment

To evaluate our hypothesis that transformations of divergent cases can facilitate the word-level alignment process, we have conducted human alignment studies for two different pairs of languages: English-Spanish and English-Arabic. We have chosen these two pairings to test the generality of the divergence transformation principle.

Our experiment involves four steps:

i. Identify 6 canonical rewritten structures—one for each divergence category.

ii. Automatically categorize English sentences into one of the 6 divergence categories (or "none") based on the foreign language.

iii. Apply the appropriate canonical rewriting to each divergence-categorized English sentence, renaming it E'.

iv. For each language:

- Human align the true English sentence and the foreign-language sentence.
- Human align the rewritten E' sentence and the foreign-language sentence.
- Compare inter-annotator agreement between the first and second sets.

First, we describe the structural representation used in our canonical rewritten structures used in i-iii above. Then we will describe our experimental setup for step iv.

5.1 Structural Representation used in our Approach

The structures used in our approach are modeled after the dependency-tree representations used in the Minipar system (Lin, 1995; Lin, 1998). We accommodate the divergence categories above by rewriting the dependency tree so that it is parallel to what would be the equivalent Spanish dependency tree. For example, consider the sentence *John kicked Mary*. Our approach rewrites the dependency tree for this sentence as a new dependency tree corresponding to the sentence *John gave kicks to Mary*.

The transformation between these two dependency tree representations is depicted in a simplified format below:

(6)(i) **Minipar Dependency Tree:**

```
(kick V_N_N root (john N subj) (mary N obj))
```

(6)(ii) **Rewritten Minipar Dependency Tree:**

```
(give V_N_N_P root (john N subj)
  (kicks N obj) (to P mod (mary N pcomp-n)))
```

Table 2 shows examples of the types of sentences that were aligned with the foreign language in our experiment (including both English and E').

5.2 Experimental Setup

In each experiment, four fluently bilingual human subjects were asked to perform word-level alignments on the same set of sentences selected from the Bible. They were all provided the same instructions and software, similar to the methodology and system described by (Melamed, 1998). Two of the four subjects were given the original English and foreign language sentences; they served as the control for the experiment. The sentence given to the other two consisted of the original foreign language

Type	English	E'	Foreign Equivalent
	fear	have fear	tiene miedo
	try	put to trying	poner a prueba
	make any cuttings	wound	تجرحوا
	our hand is high	our hand heightened	يدنا قد عظمت
	he is not here	he be-not here	إنه ليس هنا
Manner	teaches	walks teaching	anda enseñando
	is spent	self goes spending	se va gastando
	he sent his brothers away	he dismissed his brothers	صرف إخوته
	spake good of	speak-good about	يثني على
	he turned again	he returned	رجع
Head Swapping[3] Thematic	walked out	Imove-out walking	salió caminando
	I am pained	me pain they	me duelen
	He loves it	to him be-loved it	le gusta
	it was on him	was he-wears-it	كان يرتديه
Categorial	i am jealous	i have jealousy	tengo celos
	I require of you	I require-of you	te pido
	(he) shall estimate	according-to (his)-estimate	حسب تقدير
	(how long shall) the land mourn	(stays) the-land mourning	الارض نائحة
	he went to	his-return to	عودته الي
Structural	after six years	after of six years	después de seis años
	and because of that in other parts	and for that in other parts	y por ello en otras partes
	I forsake thee	I-forsake about-you	أتخلى عنك
	we found water	we-found on-water	عثرنا على

Table 2: Examples of True English, E', and Foreign Equivalent

sentences paired with altered English (denoted as E') resulting from divergence transformations described above. We compare the inter-annotator agreement rates and other relevant statistics between the two sets of human subjects. If the divergence transformations had successfully modified English structures to match those of the foreign language, we would expect the inter-annotator agreement rate between the subjects aligning the E' set to be higher than the control set. We would also expect that the E' set would have fewer unaligned and multiply-aligned words.

5.2.1 Experiment 1: English and Spanish

For the first experiment, the subjects were presented with 150 English-Spanish sentence pairs from the English and Spanish Bibles. The sentence selection procedure is similar to the divergence detection process described in the previous section. These sentences were first selected as potential divergences, using the handcrafted regular expressions referred to in Section 4; they were subsequently verified by the experimenter as belonging to a particular divergence type. The average length of the English sentences is 25.6 words; the average length of the Spanish sentences is 24.7 words. Of the four human subjects, two are native Spanish speakers, and two are Spanish literature concentrators. The backgrounds of the four human subjects are summarized in Table 3.

5.2.2 Experiment 2: English and Arabic

In the second experiment, the subjects are presented with 50 English-Arabic sentence pairs selected from the English and Arabic Bibles. While the total number of sentences is smaller than the previous experiment, many sentences contain multiple divergences. The average English sentence length is 30.5 words, and the average Arabic sentence length is 17.4 words. The backgrounds of the four human subjects are summarized in Table 4.

Inter-annotator agreement rate is quantified for each pair of subjects who viewed the same set of data. We hold one subject's alignments as the "ideal" and compute the precision and recall figures for the other subject based on how many alignment links were made by both people. The averaged precision and recall figures (F-scores)[4] for the the two experiments and other relevant statistics are summarized in Table 5. In both experiments, the inter-annotator agreement is higher for the bitext in which the divergent portions of the English sentences have been transformed. For the English-Spanish experiment, the agreement rate increased from 79.3% to 82.1%, resulting in an error reduction of 13.5%; for the English-Arabic experiment, the agreement rate increased from 69.5% to 72.5%, an error reduction of 9.8%.

[3] Although cases of Head Swapping arise in Arabic (as shown in Section 3.3), we did not find any such cases in the small sample of sentences that we human checked in the Arabic Bible.

[4] $F = \frac{2 \times Precision \times Recall}{Precision + Recall}$

	data set	native-tongue	linguistic knowledge?	ease with computers
Subject 1	control	Spanish	yes	high
Subject 2	control	Spanish	no	low
Subject 3	divergence	English	no	high
Subject 4	divergence	English	no	low

Table 3: A summary of the backgrounds of the English-Spanish subjects

	data set	native-tongue	linguistic knowledge?	ease with computers
Subject 1	control	Arabic	yes	high
Subject 2	control	Arabic	no	high
Subject 3	divergence	Arabic	no	high
Subject 4	divergence	Arabic	no	high

Table 4: A summary of the backgrounds of the English-Arabic subjects

Additional statistics also support our hypothesis that transforming divergent English sentences facilitates word-level alignment by reducing the number of unaligned and multiply-aligned words. In the English-Spanish experiment, both the appearances of unaligned words and multiply-aligned words decreased when aligning to the modified English sentences. The percentage of unaligned words decreased from 17% to 14%, and the average number of links to a word is lowered from 1.35 to 1.16.[5] In the English-Arabic experiment, the number of unaligned words is significantly smaller when aligning Arabic sentences to the modified English sentences; however, on average multiple-alignment increased. This may be due to the big difference in sentence lengths (English sentences are typically twice as long as the Arabic ones); thus it is not surprising that the average number of alignments per word would be closer to two when most of the words are aligned. The reason for the lower number in the unmodified English case might be that the subjects only aligned words that had clear translations.

6 Conclusion and Future Work

In this paper, we have described six divergence types that frequently occur in many language pairs, such as English-Spanish and English-Arabic. By examining bitext corpora, we have established conservative lower-bounds, estimating that these divergences occur at least 10% of the time. A realistic sampling indicates that the percentage is actually significantly higher, approximately 35% in Spanish.

We have shown that divergence cases can be systematically handled by transforming the syntactic structures of the English sentences to bear a closer resemblance to those of the foreign language according to a small set of templates. The validity of the divergence handling has been verified through two word-level alignment experiments. In both cases, the human subjects consistently had higher agreement rate with each other on the task of performing word-level alignment when divergent English phrases were transformed. This result suggests that divergence handling will significantly improve automatic methods of word-level alignment and facilitate cross-language processing research such as creating foreign language treebanks from projected English syntactic structures.

[5]The relatively high overall percentage of unaligned words is due to the fact that the subjects did not align punctuations.

Acknowledgements

This work has been supported, in part, by ONR MURI Contract FCPO.810548265, DARPA TIDES Contract N66001-00-2-8910, and DOD Contract MDA904-96-C-1250.

	(F-score)	% of unaligned words	Avg. alignments per word
E-S	79.3	17.2	1.35
E'-S	82.1	14.0	1.16
E-A	69.5	38.5	1.48
E'-A	72.5	11.9	1.72

Table 5: The results of the two experiments. Note: in computing the average number of alignments per word, we do not include unaligned words.

References

H. Alshawi and S. Douglas. 2000. Learning Dependency Transduction Models from Unannotated Examples. *Philosophical Transactions, Series A: Mathematical, Physical and Engineering Sciences.*

Hiyan Alshawi, Srinivas Bangalore, and Shona Douglas. 2000. Learning Dependency Translation Models as Collections of Finite State Head Transducers. *Computational Linguistics*, 26.

Jan W. Amtrup, Karine Megerdoomian, and Remi Zajac. 1999. Rapid Development of Translation Tools. In *Proceedings of the Machine Translation Summit VII*, pages 385–389, Singapore.

Peter F. Brown, John Cocke, Stephen Della-Pietra, Vincent J. Della-Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79–85, June.

Peter F. Brown, Stephen A. DellaPietra, Vincent J. DellaPietra, and Robert L. Mercer. 1993. The Mathematics of Machine Translation: Parameter Estimation. *Computational Linguistics.*

Bonnie J. Dorr. 1993. *Machine Translation: A View from the Lexicon.* The MIT Press, Cambridge, MA.

Christiane Fellbaum, Martha Palmer, Hoa Trang Dang, Lauren Delfs, and Susanne Wolff. 2001. Manual and Automatic Semantic Annotation with WordNet. In *Proceedings of the NAACL Workshop on WordNet and Other Lexical Resources: Applications, Customizations*, Carnegie Mellon University, Pittsburg, PA.

Ulf Hermjakob and Raymond J. Mooney. 1997. Learning Parse and Translation Decisions from Examples with Rich Context. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–489.

Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. 2002. Evaluating Translational Correspondence using Annotation Projection. In *submitted to the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, PA.

Rebecca Hwa. 2000. Sample selection for statistical grammar induction. In *Proceedings of the 2000 Joint SIGDAT Conference on EMNLP and VLC*, pages 45–52, Hong Kong, China, October.

Dekang Lin. 1995. Government-Binding Theory and Principle-Based Parsing. Technical report, University of Maryland. Submitted to Computational Linguistics.

Dekang Lin. 1998. Dependency-Based Evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems, First International Conference on Language Resources and Evaluation*, Granada, Spain, May.

I. Dan Melamed. 1998. Empirical Methods for MT Lexicon Development. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 18–30, Langhorne, PA, October 28–31.

Dekai Wu. 1997. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377–400.

Kenji Yamada and Kevin Knight. 2001. A Syntax-Based Statistical Translation Model. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 523–529, Toulouse, France.

David Yarowsky and Grace Ngai. 2001. Inducing Multilingual POS Taggers and NP Bracketers via Robust Projection across Aligned Corpora. In *Proc. of NAACL-2001*, pages 200–207.