Breaking the Resource Bottleneck for Multilingual Parsing

Rebecca Hwa[1], **Philip Resnik**[1,2], **and Amy Weinberg**[1,2]

Institute for Advanced Computer Studies[1]
Department of Linguistics[2]
University of Maryland, College Park, MD 20742
{hwa, resnik, weinberg}@umiacs.umd.edu

Abstract

We propose a framework that enables the acquisition of annotation-heavy resources such as syntactic dependency tree corpora for low-resource languages by importing linguistic annotations from high-quality English resources. We present a large-scale experiment showing that Chinese dependency trees can be induced by using an English parser, a word alignment package, and a large corpus of sentence-aligned bilingual text. As a part of the experiment, we evaluate the quality of a Chinese parser trained on the induced dependency treebank. We find that a parser trained in this manner out-performs some simple baselines inspite of the noise in the induced treebank. The results suggest that projecting syntactic structures from English is a viable option for acquiring annotated syntactic structures quickly and cheaply. We expect the quality of the induced treebank to improve when more sophisticated filtering and error-correction techniques are applied.

1 Introduction

There is a substantial disparity between the quality of state of the art parsers available for English and those for other languages. English parsers such as those of Collins (1997) and Charniak (1999) were trained on hand annotated corpora such as the Penn Treebank Project (Marcus et al., 1993). However, experience has shown us that building hand-crafted treebanks from scratch is too time-consuming to be repeated for every language of interest. This bad news can be mitigated by leveraging English annotations to automatically acquired annotations for new languages. Recent work by Yarowsky and Ngai (2001) has shown that this type of transfer is possible for inducing part-of-speech tags for Chinese. In this paper, we explore the application of this technique to the more complex problem of inducing Chinese dependency trees.

The input to our system is a collection of sentence-aligned bilingual text (i.e., pairs of sentences that are translations of each other). Each English sentence is parsed using a high-quality English parser. For each pair of sentences, word alignment is performed using statistical MT models (Brown et al., 1990; Al-Onaizan et al., 1999). The alignment then anchors the projection of the English tree to the Chinese side (see Figure 1).

This paper presents an initial large-scale experiment, investigating the feasibility of inducing a Chinese dependency treebank using our projection algorithm and of training a parser on the resulting treebank. Due to the compounded errors of various components of the system, the induced Chinese dependency treebank is rather noisy. Applying filtering heuristics to the treebank improves its quality enough such that the parser trained on it out-performs some

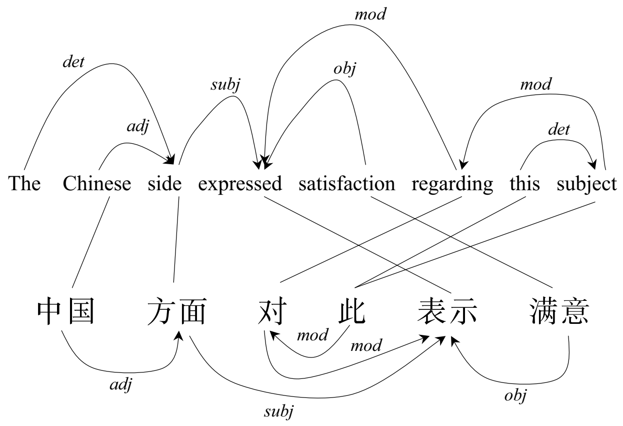

Figure 1: Given an English dependency parse tree and a set of word alignments, we infer the syntactic structure on the Chinese side via projection from its English counterpart.

simple baselines. While the parser's performance is still significantly less than that of a parser trained on a clean, fully annotated (Chinese) treebank, this study suggests that projecting syntactic structures from English is viable for acquiring annotated syntactic structures quickly and cheaply.

2 Overview of the Algorithm

Our approach requires three resources. First, we need a sizable, sentence-aligned bilingual text as training corpus. In our experiment, we use a bilingual text of English and Chinese news articles. In Section 5 we discuss other ways in which bilingual text can be acquired and sentence aligned. Second, we require dependency parses of the English text. Our choice of dependency representation is motivated in Section 2.1. Third, word alignments are needed to relate the sentence pair on the lexical level. In this paper, we use alignments produced as a side-effect of training a statistical translation model (Brown et al., 1990; Al-Onaizan et al., 1999).

Given these resources, our system behaves as follows: for each sentence pair (E, C) in the bilingual text, the English sentence E is parsed and converted into a dependency representation. Next, word alignment is performed for the sentence pair. Finally, the English dependency analysis is projected across the word alignment to the Chinese side according to our *Direct Projection Algorithm*, which we outline in section 2.2.

2.1 Dependency Representations as Transfer Medium

Dependency relationships specify asymmetric binary relations between two surface words: a *head* and its *modifier*. For example, in the sentence from Figure 1, *"The Chinese side expressed satisfaction regarding this subject,"* the word *side* modifies the head word *expressed*. The dependency links may optionally be annotated with information specifying grammatical relations between constituents such as *subject*, *object*, *modifier*, etc. In our example, the link between *side* and *expressed* is labeled as *subj*, indicating that the constituent *The Chinese side* is the subject of the verb *expressed*. In this section, we argue that dependency representation is right for our projection framework because it captures both structural and lexical relationships between words that are not string local; because it overcomes some of the shortcomings of evaluating against the phrase structure representation; and because it is language independent with respect to word order variations.

Syntactic analysis in terms of phrase structure has been the dominant paradigm in natural language processing, starting from early context-free grammars and continuing up to present-day stochastic formalisms. It is preferable over models that make Markov assumptions restricting interactions among words to those that occur within the window of an n-gram. Phrase structure formalisms provide a level of representation that allows significant constraint to occur between grammatical categories that are not string-local. These categories become *local* at the phrase structure level. For example, consider the following sentence from the Brown Corpus:

> *The largest hurdle the Republicans would have to face is a state law which says that before making a first race, one of two alternative courses must be taken.*

The relationship between *hurdle* and *is* exists over a long string-distance, owing to an embedded relative clause, and, similarly, *Republicans* and *face* are separated in the string by a sequence of auxiliaries and the infinitival *to*. As a result, the relationships represented in the sentence are not captured well by any n-gram model with tractable n. In contrast, the relationship between the subject NP and the predicate is easily encoded locally within a context-free rule such as S $\rightarrow$ NP VP.

To take full advantage of such relationships in models based on phrase structure, however, it is necessary to *lexicalize* the grammar formalism, so that lexically-based constraints are also localized within grammar rules. By incorporating lexical content into phrase structure rules (e.g., S $\rightarrow$ NP(hurdle) VP(is)), lexicalized grammar formalisms make it possible to capture syntactic constraints such as as number agreement (e.g. the low probability of S $\rightarrow$ NP(hurdle) VP(are)) as well as semantic constraints (e.g. the reasonably high probability of S $\rightarrow$ NP(Republicans) VP(face)). Work taking advantage of this insight (e.g. Collins (1997; Charniak (1999)) has defined the breakthroughs leading to the current state of the art in broad-coverage parsing. Implicitly or sometimes explicitly (as in the work of

Collins), what gives lexicalized context-free representations their power is the ability to probabilistically model the syntactic *dependency* relationships between words in the structure.

Moreover, dependency analysis evaluation avoids some of the shortcomings of constituency analysis evaluation (Lin, 1995; Carroll et al., 1999). Standard constituency parsing metrics compare the phrase boundaries specified by the gold standard to that of the candidate analysis. They also evaluate whether conditions on well formed trees (such as a ban on crossing branches) are respected by the candidate. However, as Lin (1995) notes, since branching structure is not directly tied to semantic interpretation, it is unclear how to interpret missing, spurious, or crossing branches. On the other hand, it is apparent that syntactic dependencies, more so than syntactic constituents, are closely tied to the who-did-what-to-whom relationships of language. Indeed, work in lexical semantics relating syntactic representations to thematic relationships such as *agent*, *theme*, *beneficiary*, has focused primarily on syntactic dependencies rather than on phrasal constituents (Baker, 1997). Since semantic dependencies form a superset based on syntactic dependencies, we are better able to gauge how likely a representation is to be interpretable, by measuring the percentage of correct dependencies.

Finally, dependency structures firmly separate precedence from dominance relations, such that word order variation between languages becomes less of a problem than in constituency trees. For example, the relative string order of a series of modifiers of a head is irrelevant in the dependency representation. All are modifiers. By contrast, a constituency tree may require a stacked structure that would not translate well if the word order were reversed in another language. In other words, dependency structures are more likely to respect a homomorphism.

These observations suggest that dependencies may be a better choice for syntactic projection across languages than phrasal constituents. To the extent that this assumption is correct, we should be able to use word alignments as a bridge between English and another language, retaining some level of confidence that if dependencies are projected across the alignment they will be correct for the new language. Experimental results from our previous work (Hwa et al., 2002), have indicated that while the assumption does not always hold true, syntactic analyses projected from English to Chinese can, in principle, yield Chinese analyses that are nearly 70% accurate (in terms of unlabeled dependencies) after application of a set of linguistically principled rules.[1]

2.2 The Direct Projection Algorithm

Our approach is based on the intuitive idea of a direct projection of dependency structures. We now describe our projection algorithm in more detail. Given sentence pair (E, C), where $E = e_1, \ldots, e_n$ and $C = c_1, \ldots, c_m$, syntactic relations (denoted as $R(x, y)$) are projected from English for the following situations:

- **one-to-one** if e_i is aligned with a unique c_x and e_j is aligned with a unique c_y, if $R(e_i, e_j)$, conclude $R(c_x, c_y)$.

- **unaligned (English)** if e_j is not aligned with any word in C, then create a new empty word c_y such that for any e_i aligned with a unique c_x, $R(e_i, e_j) \Rightarrow R(c_x, c_y)$ and $R(e_j, e_i) \Rightarrow R(c_y, c_x)$.

- **one-to-many** if $e_i$is aligned with $c_x, \ldots, c_y$, then create a new empty word c_z such that c_z is the parent of $c_x, \ldots, c_y$ and set e_i to align to c_z instead. We called this a *Multiply-Aligned Component*, or MAC.

- **many-to-one** if $e_i, \ldots, e_j$ are all uniquely aligned to c_x, then delete all alignments between $e_k (i \leq k \leq j)$ and c_x except for the head of $e_i, \ldots, e_j$.

The **many-to-many** case is decomposed into a two-step process: first perform one-to-many, then perform many-to-one. In the cases of unaligned Chinese words, they are left out of the projected syntactic tree. The asymmetry of the treatment of **one-to-many** and **many-to-one** and of the unaligned words for the two languages arises from the asymmetric nature of the projection.

[1]The experiment was performed under idealized settings, projecting human annotated English dependency analyses using human annotated word alignments.

2.2.1 Post-Projection Transformation

The Direct Projection Algorithm by itself does not produce good dependency trees because it does not properly handle structural projection for the more complex cases when the alignment is not one-to-one. Therefore, we apply a small set of linguistically motivated rules to correct the projected trees as a post-hoc process. It is clearly an advantage to limit the correction rules to those that can apply generally, across many construction types. Wanting to avoid unending language-specific rule tweaking, we strictly limited the possible rules. Rules were permitted to refer only to closed class items, to parts of speech projected from the English analysis, or to easily enumerated lexical categories (e.g. $\{$*dollar*, RMB, \$, yen$\}$). The majority of rule patterns are variations on the same solution to the same problem. Viewing the problem from a higher level of linguistic abstraction made it possible to find all the relevant cases in a short time and express the solution compactly; in all, fewer than twenty rules were written, and the analysis, rule writing, and verification of their correctness using the data set took a few days.

Here are two examples of the rules we developed; see (Hwa et al., 2002) for fuller discussion.

Rule for noun modification:

- If $c_x, \ldots, c_y$ are a set of Chinese words aligned to an English noun, replace the empty node introduced in the Direct Projection Algorithm by promoting the last word c_y to its place with $c_x, \ldots, c_{y-1}$ as dependents.

Rule for aspectual markers:

- If $c_x, \ldots, c_y$, a sequence of Chinese words aligned with English verbs, is followed by c_a, an aspect marker, make c_a into a modifier of the last verb c_y.

2.2.2 Remaining Shortcomings of the Direct Projection Algorithm

Although the majority of the projected trees are significantly improved, the post-projection transformation rules still do not adequately address some major deficiencies of the Direct Projection Algorithm. The algorithm does not ensure that the projected structure is indeed a well-formed structure. Thus, when given unconstrained word alignment outputs, the projected structure may contain errors such as crossing dependencies (see Figure 2). Moreover, due to the asymmetry of the algorithm, the syntactic role of unaligned foreign words cannot be inferred. The post-projection transformation rules address this problem to some extent by incorporating unaligned function words back into the parse, but an intelligent treatment of the open class of unaligned words remains a challenge of this projection approach. Furthermore, the algorithm does not address complex translation divergences (Dorr, 1993), such as the head-swapping phenomenon (in which the direction of the head-modifier dependency is reversed in the foreign language). Lopez et al. (2002) describe an alternative to the direct projection approach that addresses some of these problems.

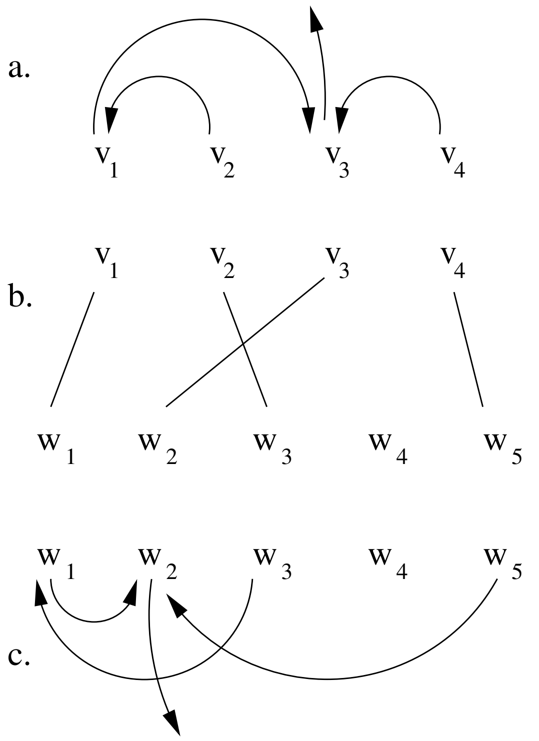

Figure 2: The direct projection of the dependency parse for $v_1 \ldots v_4$ (Figure 2a) across the word alignment (Figure 2b) results in cross dependency relationships for the link between w_1 and w_3 and the link between w_2 and w_5; and it leaves word w_4 unattached to the projected dependency tree (Figure 2c).

3 Experimental Setup

Our previous results have shown that, given good English parses and clean alignments to Chinese translations, the direct projection approach from English to Chinese (together with post-processing) can lead to Chinese annotations that are substantially correct; unlabeled precision/recall on projected dependencies approaches 70% (Hwa et al., 2002). While this demonstrates that the approach holds promise in automatically inducing syntactic treebanks of reasonable quality, it is not clear how much degradation occurs when using imperfect English parsers and imperfect word alignment models. That question is our focus in this paper. We report a full-scale experiment on English and Chinese sentence pairs, evaluating the entire framework under the realistic settings of imperfect bilingual data and error-prone parsers and alignment models (see Section 3.1). Once a Chinese dependency treebank is induced, we use it to train a Chinese parser in a manner similar to that of Collins (1999). The trained parser is then evaluated on unseen test sentences taken from the Chinese Treebank (Xia et al., 2000) and compared with two baselines and an upper bound.

3.1 Resources

We use about 56,000 sentence pairs from the Hong Kong News (HKNews) corpus as our bilingual text. The data have been automatically sentence aligned and the Chinese words have been automatically segmented.[2] To parse the English sentences, we use a lexicalized statistical parser trained on the Wall Street Journal corpus (Collins, 1997).[3] To obtain word alignments for all sentence pairs, we train an off-the-shelf statistical translation model, GIZA++ (Al-Onaizan et al., 1999), using the HKNews bilingual text. Given these resources, the direction projection algorithm and the post-projection transformation process are then used to induce dependency trees for the Chinese sentences in the HKNews corpus.

3.2 Evaluation of the Induced Treebank

Because of its size, we do not directly assess the quality of the induced treebank. Instead, we evaluate the Chinese parser trained from it. To the extent that the trained parser outputs reasonable structures on unseen test sentences, it indicates that the induced treebank is a useful resource. To evaluate the quality of the trained parser, we compare it to two simple baseline dependency analyses: *always modify the previous word*, and *always modify the next word*. As an upper bound, we have also trained the same parser with clean, hand-annotated trees from the Penn Chinese Treebank (ChTB). We constructed a development set consisting of 124 sentences and a test set consisting of 88 sentences taken from the Chinese Treebank; all sentences are of 40 words or less. The remaining approximately 3800 Chinese Treebank sentences are converted into their dependency representation (similar to the algorithm described in Section 2 of the paper by Xia and Palmer (2001)) and used as training data for the upper-bound parser. We evaluate the trained parser by comparing its output (dependency) parse trees for the unseen test sentences against the human-annotated gold standard parse trees (also converted to dependency representation). The metrics used are the precision and recall scores on the unlabeled dependency relations. A parser produced dependency link is considered "correct" if the same head-modifier relationship exists in the gold standard; the dependency label does not need to match. Punctuations are not scored.

4 Results and Discussions

Tables 1 and 2 show performance comparisons for our automatic projection approach as compared to the lower and upper bounds. As one might expect, the quality of the treebank induced under the real-world constraints of imperfect data and components is noticeably worse than one induced using clean English parses and perfect word alignments. The Direct Projection Algorithm and its associated post-projection transformation rules are not fault-tolerant enough to recover from the compounding errors of the parser and alignment model. Without further processing, the projected treebank would contain too much noise to be useful for training a parser. Therefore, our attentions turn to filtering heuristics for poorly induced dependency trees.

We found that the most unreliable component is the word alignment model. A cursory inspection of the alignment output (for the HKNews corpus) shows

[2] We are grateful to Stefan Vogel of CMU for his assistance with this corpus.

[3] The executable of the parser is freely available at ftp://ftp.cis.upenn.edu/pub/mcollins/misc.

that, for many sentences, the majority of the English words remain unaligned; and that often, an unusually high number of Chinese words (e.g, five or greater) are aligned to the same English word. The poor alignment output may have many causes: in particular, the sentence pair input to the alignment model is imperfect, and the alignment model does not perform well for language pairs with very dissimilar word-order patterns.

This suggests that performance might improve if we filter out sentence pairs that are known to be poorly aligned. To filter out dependency trees projected from dubious word alignments, we have devised several simple heuristics. First, we removed those sentences for which more than 30% of the English words were not aligned to any Chinese word ($EnoC \leq 0.3$). The figure 30% is empirically determined, based on the trained parser's performance on the development set. As shown in the first row of Table 1, the parser trained on the filtered treebank does outperform the modify-next baseline; however, the corpus size has been drastically cut-down from around 56,000 to less than 8,000. The second filter we apply to the corpus is to remove sentences in which the size of a multiply aligned component is greater than three (MAC $>$ 3); that is, when more than three Chinese words are aligned to the same English word. The MAC value of 3 was also determined empirically using development data. The second line of Table 1 shows that training the parser on the induced treebank filtered by both heuristics leads to further improvement. Finally, we return to the crossing-dependency problem alluded to earlier in section 2.2.2. While we do not correct the crossing dependencies in this work, we remove sentences with the most egregious crossing-dependency violations in their analyses. Our experiments with development data suggested that a sentence should be filtered out if more than 40% of its dependency links violate the no-crossing constraint. The combination of the three filters improved the induced treebank so that a parser trained on the treebank outperforms the simple baselines; however, the draconian filters also reduced the corpus from 56,000 sentences to slightly over 5,000.

Table 2 shows the trained parser's performance on a separate test set. As before, it is compared with two baselines; and as an upper bound, we train the same parser on a clean, manually created treebank.[4] Similar to the outcome of the development set, the trained parser performs better than the baseline, but it still cannot compete with a parser trained on a clean corpus. It is interesting to note that after our current filtering techniques, the sizes of the induced treebank is comparable to the clean one. However, our method of treebank acquisition is not constrained by the laborious manual annotation process; therefore it would be easy for us to obtain a much larger bilingual corpus as a starting point, as discussed below. We conjecture that the size of the corpus will help offset the effect of the noise, as will more sophisticated sampling techniques that exclude the noisiest data.

5 Conclusion and Future Work

In this paper, we have described our framework for acquiring Chinese dependency treebanks by bootstrapping from existing linguistic resources for English. We have explicitly discussed the assumptions made and the resources required in order for our algorithm to work. An ambitious full-scale experiment using real-world data was performed to investigate the feasibility of our approach. Our results suggest that treebank acquisition through projection is indeed possible; however reducing the noise in the induced treebank is a major challenge.

This finding points us to several directions for further research. One clear avenue is to obtain larger bilingual texts, so that more data remain even when noisy sentence pairs have been filtered out. Work on mining the Web for bilingual text, such as STRAND (Resnik, 1999), BITS (Ma and Liberman, 1999), and PTMiner (Nie et al., 1999), show significant promise in this regard. Once parallel Web pages are obtained, it is possible to obtain sentence- or segment-level alignments either via alignment of HTML markup (Resnik, 1998) or via more sophisticated sentence-alignment techniques (Melamed, 1998).

Beyond simply taking a "more is better" approach to data acquisition, one way to reduce the noise in the induced treebank is to lower the error rates of the individual components in our projection framework. Of these, improving the word alignment model

[4] The upper-bound parser's performance is on par with that of the state of the art constituency parsers trained on the Chinese Treebank, e.g. (Bikel and Chiang, 2000).

Method	Corpus Size	Precision & Recall
EnoC	7689	37.4
EnoC+MAC	5525	42.1
EnoC+MAC+NoCross	5284	42.9
Modify Prev (Baseline)	–	14.0
Modify Next (Baseline)	–	32.2

Table 1: The parser's performance on the development set (%) when the training corpus has been filtered with the following heuristics: remove sentences if too many English words have no Chinese translations (EnoC); remove sentences if too many Chinese words are aligned to one English word (MAC); remove sentences that violate many crossing-dependency constraints (NoCross).

Method	Corpus	Size	Precision & Recall
Modify Prev (Baseline)	–	–	13.5
Modify Next (Baseline)	–	–	35.7
Stat. Parser	Induced HKNews	5284	42.3
Stat. Parser (Upper-bound)	Clean ChTB	3870	75.6

Table 2: A comparison of the parsers' performance against lower and upper bounds on the test set (%).

would benefit the overall system the most. We are actively developing alternative word alignment models that is sensitive to this syntactic projection framework (Lopez et al., 2002). Moreover, as we have shown in this study, filtering techniques that identify and remove malformed trees can help reducing noise; however, aggressive filtering alone is likely to result in over-filtering. To render nearly 90% of the bilingual text useless places too heavy a burden on even the best Web mining techniques. We are experimenting with filtering strategies that attempt to localize the potentially problematic parts of a syntactic tree so that the rest can still contribute to the training corpus. In addition, we are continuing to work on the post-projection transformation the process to improve the quality of the projected trees.

6 Acknowledgments

This work has been supported, in part, by ONR MURI Contract FCPO.810548265, NSA RD-02-5700, DARPA/ITO Cooperative Agreement N660010028910 and Mitre Contract 010418-7712. The authors would like to thank Franz Josef Och for his help with using the GIZA++ translation model; and Adam Lopez and Mike Nossal for helpful discussions and comments on this paper.

7 References

Yaser Al-Onaizan, Jan Curin, Michael Jahr, Kevin Knight, John Lafferty, I. Dan Melamed, Franz-Josef Och, David Purdy, Noah A. Smith, and David Yarowsky. 1999. Statistical machine translation. Technical report, JHU. citeseer.nj.nec.com/al-onaizan99statistical.html.

Mark C. Baker, 1997. *Thematic Roles and Syntactic Structure*, pages 73–137. Kluwer.

Daniel Bikel and David Chiang. 2000. Two statistical parsing models applied to the chinese treebank. In *Proceedings of the Second Chinese Language Processing Workshop*, pages 1–6.

Peter F. Brown, John Cocke, Stephen A. DellaPietra, Vincent J. DellaPietra, Frederick 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.

John Carroll, Guido Minnen, and Ted Briscoe. 1999. Corpus annotation for parser evaluation. In *LINC-99 workshop at the 9th Conference of the EACL*, June.

Eugene Charniak. 1999. A maximum-entropy inspired parser. Technical Report CS-99-12, Brown University.

Michael Collins. 1997. Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the ACL*, pages 16–23, Madrid, Spain.

Michael Collins. 1999. A statistical parser for czech. In *Proceedings of the 37th Annual Meeting of the ACL*, College Park, Maryland.

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

Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. 2002. Evaluating translational correspondence using annotation projection. In *Proceedings of the 40th Annual Meeting of the ACL*. To appear.

Dekang Lin. 1995. A dependency-based method for evaluating broad-coverage parsers. In *Proceedings of the IJCAI-95*, pages 1420–1425.

Adam Lopez, Michael Nossal, Rebecca Hwa, and Philip Resnik. 2002. Word-level alignment for multilingual resource acquisition. In *Proceedings of the Workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*. To appear.

Xiaoyi Ma and Mark Liberman. 1999. Bits: A method for bilingual text search over the web. In *Machine Translation Summit VII*.

Mitchell Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19(2):313–330.

I. Dan Melamed. 1998. *Empirical Methods for Exploiting Parallel Texts*. Ph.D. thesis, University of Pennsylvania, May.

J. Nie, M. Simard, P. Isabelle, and R. Durand. 1999. Cross-language information retrieval based on parallel texts and automatic mining parallel texts from the web. In *Proceedings of the ACM SIGIR Conference*.

Philip Resnik. 1998. Parallel strands: A preliminary investigation into mining the Web for bilingual text. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, Langhorne, PA, October 28-31.

Philip Resnik. 1999. Mining the web for bilingual text. In *Proceedings of the 37th Annual Meeting of the ACL*, June.

Fei Xia and Martha Palmer. 2001. Converting dependency structures to phrase structures. In *Proc. of the HLT Conference*, March.

Fei Xia, Martha Palmer, Nianwen Xue, Mary Ellen Ocurowski, John Kovarik, Fu-Dong Chiou, Shizhe Huang, Tony Kroch, and Mitch Marcus. 2000. Developing guidelines and ensuring consistency for chinese text annotation. In *Proceedings of the Second Language Resources and Evaluation Conference*, June.

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.