Evaluating Translational Correspondence using Annotation Projection

Rebecca Hwa[1], **Philip Resnik**[1,2], **Amy Weinberg**[1,2], **and Okan Kolak**[1,3]
Institute for Advanced Computer Studies[1]
Department of Linguistics[2]
Department of Computer Science[3]
University of Maryland, College Park, MD 20742
{hwa, resnik, weinberg, okan}@umiacs.umd.edu

Abstract

Recently, statistical machine translation models have begun to take advantage of higher level linguistic structures such as syntactic dependencies. Underlying these models is an assumption about the directness of translational correspondence between sentences in the two languages; however, the extent to which this assumption is valid and useful is not well understood. In this paper, we present an empirical study that quantifies the degree to which syntactic dependencies are preserved when parses are projected directly from English to Chinese. Our results show that although the direct correspondence assumption is often too restrictive, a small set of principled, elementary linguistic transformations can boost the quality of the projected Chinese parses by 76% relative to the unimproved baseline.

1 Introduction

Advances in statistical parsing and language modeling have shown the importance of modeling grammatical dependencies (i.e., relationships between syntactic heads and their modifiers) between words (Collins, 1997; Eisner, 1997; Chelba and Jelinek, 1998; Charniak, 2001). Informed by the insights of this work, recent statistical machine translation (MT) models have become linguistically richer in their representation of monolingual relationships than their predecessors ((Wu, 1995; Alshawi et al., 2000; Yamada and Knight, 2001); cf. (Brown et al., 1990; Brown et al., 1993)).

Using richer monolingual representations in statistical MT raises the challenge of how to characterize the cross-language relationship between two sets of monolingual syntactic relations. In this paper, we investigate a characterization that often appears implicitly as a part of newer statistical MT models, which we term the *direct correspondence assumption (DCA)*. Intuitively, the assumption is that for two sentences in parallel translation, the syntactic relationships in one language directly map to the syntactic relationships in the other. Since it has not been described explicitly, the validity and utility of the DCA are not well understood — although, without identifying the DCA as such, other translation researchers have nonetheless found themselves working around its limitations.[1]

In Section 2 we show how the DCA appears implicitly in several models, providing an explicit formal statement, and we discuss its potential inadequacies. In Section 3, we provide a way to assess empirically the extent to which the DCA holds true. Our results suggest that although the DCA is too restrictive in many cases, a general set of principled, elementary linguistic transformations can often resolve the problem.

[1]For example, Yamada and Knight (2001) account for non-DCA-respecting variation by learning construction specific local transformations on constituency trees. There also exists a substantial literature in transfer-based MT on learning mapping patterns for syntactic relationships that do not correspond (e.g., (Menezes and Richardson, 2001; Lavoie et al., 2001)).

In Section 4, we consider the implications of our experimental results and discuss future work.

2 The Direct Correspondence Assumption

To our knowledge, the direct correspondence assumption underlies all statistical models that attempt to capture a relationship between syntactic structures in two languages, be they constituent models or dependency models. As an example of the former, consider Wu's (1995) stochastic inversion transduction grammar (SITG), in which paired sentences are simultaneously generated using context-free rules; word order differences are accounted for by allowing each rule to be read in a left-to-right or right-to-left fashion, depending on the language. For example, SITG can generate verb initial (English) and verb final (Japanese) verb phrases using the same rule VP $\rightarrow$ V NP. For any derivation using this rule, if V_E and NP_E are the English verb and noun phrase, and they are respectively aligned with Japanese verb and noun phrase V_J and NP_J, then VERB-OBJECT($\mathrm{V}_E, \mathrm{NP}_E$) and VERB-OBJECT($\mathrm{V}_J, \mathrm{NP}_J$) must both be true.

As an example where the DCA relates dependency structures, consider the hierarchical alignment algorithm proposed by Alshawi et al. (2000). In this framework, word-level alignments and paired dependency structures are constructed simultaneously. The English-Basque example (1) illustrates: if the English word *buy* is aligned to the Basque word *erosi* and *gift* is aligned to *opari*, the creation of the head-modifier relationship between *buy* and *gift* is accompanied by the creation of a corresponding head-modifier relationship between *erosi* and *opari*.

(1) a. I got a gift for my brother
b. Nik (I) nire (MY) anaiari (BROTHER-DAT) opari (GIFT) bat (A) erosi (BUY) nion (PAST)

2.1 Formalizing the DCA

Let us formalize this intuitive idea about corresponding syntactic relationships in the following more general way:

Direct Correspondence Assumption (DCA): Given a pair of sentences E and F that are (literal) translations of each other with syntactic structures $Tree_E$ and $Tree_F$, if nodes x_E and y_E of $Tree_E$ are aligned with nodes x_F and y_F of $Tree_F$, respectively, and if syntactic relationship $R(x_E, y_E)$ holds in $Tree_E$, then $R(x_F, y_F)$ holds in $Tree_F$.

Here, $R(x, y)$ may specify a head-modifier relationship between words in a dependency tree, or a sisterhood relationship between non-terminals in a constituency tree. As stated, the DCA amounts to an assumption that the cross-language alignment resembles a homomorphism relating the syntactic graph of E to the syntactic graph of F.[2]

Wu's SITG makes this assumption, under the interpretation that R is the head-modifier relation expressed in a rewrite rule. The IBM MT models (Brown et al., 1993) do not respect the DCA, but neither do they attempt to model any higher level syntactic relationship between constituents within or across languages—the translation model (alignments) and the language model are statistically independent. In Yamada and Knight's (2001) extension of the IBM models, on the other hand, grammatical information from the source language is propagated into the noisy channel, and the grammatical transformations in their channel model appear to respect direct correspondence.[3] The simultaneous parsing and alignment algorithm of Alshawi et al. (2000) is essentially an implementation of the DCA in which relationship R has no linguistic import (i.e. anything can be a head).

[2] Some models embody a stronger version of the DCA that more closely resembles an isomorphism between dependency graphs(Shieber, 1994), though we will not pursue this idea further here.

[3] Knight and Yamada actually pre-process the English input in cases that most transparently violate direct correspondence; for example, they permute English verbs to sentence-final position in the model transforming English into Japanese. Most models we looked at have addressed some effects of DCA failure, but they have not acknowledged it explicitly as an underlying assumption, nor have they gone beyond expedient measures to the type of principled analysis that we propose below.

R	x_{Eng}	y_{Eng}	x_{Bsq}	y_{Bsq}
verb-subj	got	I	erosi	nik
verb-obj	got	gift	erosi	opari
noun-det	gift	a	opari	bat
noun-mod	brother	my	anaiari	nire

Table 1: Correspondences preserved in (1)

2.2 Problems with the DCA

The DCA seems to be a reasonable principle, especially when expressed in terms of syntactic dependencies that abstract away word order. That is, the thematic (who-did-what-to-whom) relationships are likely to hold true across translations even for typologically different languages. Consider example (1) again: despite the fact that the Basque sentence has a different word order, with the verb appearing at the far right of the sentence, the syntactic dependency relationships of English (subject, object, noun modifier, etc.) are largely preserved across the alignment, as illustrated in Table 1. Moreover, the DCA makes possible more elegant formalisms (e.g. SITG) and more efficient algorithms. It may allow us to use the syntactic analysis for one language to infer annotations for the corresponding sentence in another language, helping to reduce the labor and expense of creating treebanks in new languages (Cabezas et al., 2001; Yarowsky and Ngai, 2001).

Unfortunately, the DCA is flawed, even for literal translations. For example, in sentence pair (1), the indirect object of the verb is expressed in English using a prepositional phrase (headed by the word *for*) that attaches to the verb, but it is expressed with the dative case marking on *anaiari* (BROTHER-DAT) in Basque. If we aligned both *for* and *brother* to *anaiari*, then a many-to-one mapping would be formed, and the DCA would be violated: $R(for, brother)$ holds in the English tree but $R(anaiari, anaiari)$ does not hold in the Basque tree. Similarly, a one-to-many mapping (e.g., aligning *got* with *erosi* (BUY) and *nion* (PAST) in this example) can also be problematic for the DCA.

The inadequacy of the DCA should come as no surprise. The syntax literature dating back to Chomsky (1981), together with a rich computational literature on translation divergences (e.g. (Abeille et al., 1990; Dorr, 1994; Han et al., 2000)), is concerned with characterizing in a systematic way the apparent diversity of mechanisms used by languages to express meanings syntactically. For example, current theories claim that languages employ stable head-complement orders across construction types. In English, the head of a phrase is uniformly to the left of modifying prepositional phrases, sentential complements, etc. In Chinese, verbal and prepositional phrases respect the English ordering but heads in the nominal system uniformly appear to the right. Systematic application of this sort of linguistic knowledge turns out to be the key in getting beyond the DCA's limitations.

3 Evaluating the DCA using Annotation Projection

Thus far, we have argued that the DCA is a useful and widely assumed principle; at the same time we have illustrated that it is incapable of accounting for some well known and fundamental linguistic facts. Yet this is not an unfamiliar situation. For years, stochastic modeling of language has depended on the linguistically implausible assumptions underlying n-gram models, hidden Markov models, context-free grammars, and the like, with remarkable success. Having made the DCA explicit, we would suggest that the right questions are: *to what extent* is it true, and *how useful* is it when it holds?

In the remainder of the paper, we focus on answering the first question empirically by considering the syntactic relationships and alignments between translated sentence pairs in two distant languages (English and Chinese). In our experimental framework, a system is given the "ideal" syntactic analyses for the English sentences and English-Chinese word-alignments, and it uses a Direct Projection Algorithm (described below) to project the English syntactic annotations directly across to the Chinese sentences in accordance with the DCA. The resulting Chinese dependency analyses are then compared with an independently derived gold standard, enabling us to determine recall and precision figures for syntactic dependencies (cf. (Lin, 1998)) and to perform a qualitative error analysis. This error analysis led us to revise our projection approach, and the resulting linguistically informed projection improved significantly the ability to obtain accurate Chinese parses.

This experimental framework for the first question is designed with an eye toward the second, concerning the usefulness of making the direct correspondence assumption. If the DCA holds true more often than not, then one might speculate that the projected syntactic structures could be useful as a treebank (albeit a noisy one) for training Chinese parsers, and could help more generally in overcoming the syntactic annotation bottleneck for languages other than English.

3.1 The Direct Projection Algorithm

The DCA translates fairly directly into an algorithm for projecting English dependency analyses across to Chinese using word alignments as the bridge. More formally, given sentence pair (E, F), the English syntactic relations are projected for the following situations:

- **one-to-one** if $h_E \in E$ is aligned with a unique $h_F \in F$ and m_E is aligned with a unique $m_F \in F$, then if $R(h_E, m_E)$, conclude $R(h_F, m_F)$.

- **unaligned (English)** if $w_E \in E$ is not aligned with any word in F, then create a new empty word $n_F \in F$ such that for any x_E aligned with a unique x_F, $R(x_E, w_E) \Rightarrow R(x_F, n_F)$ and $R(w_E, x_E) \Rightarrow R(n_F, x_F)$.

- **one-to-many** if $w_E \in E$ is aligned with $w_{1_F}, \ldots, w_{n_F}$, then create a new empty word $m_F \in F$ such that m_F is the parent of $w_{1_F}, \ldots, w_{n_F}$ and set w_E to align to m_F instead.

- **many-to-one** if $w_{1_E}, \ldots, w_{n_E} \in E$ are all uniquely aligned to $w_F \in F$, then delete all alignments between $w_{i_E} (1 \leq i \leq n)$ and w_F except for the head (denoted as w_{h_E}); moreover, if w_{i_E}, a modifier of w_{h_E}, had its own modifiers, $R(w_{i_E}, w_{j_E}) \Rightarrow R(w_{h_F}, w_{j_F})$.

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 in 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.

3.2 Experimental Setup

The corpus for this experiment was constructed by obtaining manual English translations for 124 Chinese newswire sentences (with 40 words or less) contained in sections 001-015 of the Penn Chinese Treebank (Xia et al., 2000). The Chinese data in our set ranged from simple sentences to some complicated constructions such as complex relative clauses, multiple run-on clauses, embeddings, nominal constructions, etc. Average sentence length was 23.7 words.

Parses for the English sentences were constructed by a process of automatic analysis followed by hand correction; output trees from a broad-coverage lexicalized English parser (Collins, 1997) were automatically converted into dependencies to be corrected. The gold-standard dependency analyses for the Chinese sentences were constructed manually by two fluent speakers of Chinese, working independently and using the Chinese Treebank's (manually constructed) constituency parses for reference.[4] Inter-annotator agreement on unlabeled syntactic dependencies is 92.4%. Manual English-Chinese alignments were constructed by two annotators who are native speakers of Chinese using a software environment similar to that described by Melamed (1998).

The direct projection of English dependencies to Chinese yielded poor results as measured by precision and recall over unlabeled syntactic dependencies: precision was 30.1% and recall 39.1%. Inspection of the results revealed that our manually aligned parallel corpus contained many instances of multiply aligned or unaligned tokens, owing either to freeness of translation

[4] One author of this paper served as one of the annotators.

(a violation of the assumption that translations are literal) or to differences in how the two languages express the same meaning. For example, to quantify a Chinese noun with a determiner, one also needs to supply a measure word in addition to the quantity. Thus, the noun phrase *an apple* is expressed as *yee* (AN) *ge* (-MEAS) *ping-guo* (APPLE). Chinese also includes separate words to indicate aspectual categories such as continued action, in contrast to verbal suffixes in English such as the *-ing* in *running*. Because Chinese classifiers, aspectual particles, and other functional words do not appear in the English sentence, there is no way for a projected English analysis to correctly account for them. As a result, the Chinese dependency trees usually fail to contain an appropriate grammatical relation for these items. Because they are frequent, the failure to properly account for them significantly hurts performance.

3.3 Revised Projection

Our error analysis led to the conclusion that the correspondence of syntactic relationships would be improved by a better handling of the **one-to-many** mappings and the **unaligned** cases. We investigated two ways of addressing this issue.

First, we adopted a simple strategy informed by the tendency of languages to have a consistent direction for "headedness". Chinese and English share the property that they are head-initial for most phrase types. Thus, if an English word aligns to multiple Chinese words $c_1, \ldots, c_n$, the leftmost word c_1 is treated as the head and $c_2, \ldots, c_n$ are analyzed as its dependents. If a Chinese empty node was introduced to align with an untranslated English word, it is deleted and its left-most child is promoted to replace it. Looking at language in this non-construction-dependent way allows us to make simple changes that have wide ranging effects. This is illustrative of how our approach tries to rein in cases where the DCA breaks down by using linguistically informed constraints that are as general as possible.

Second, we used more detailed linguistic knowledge of Chinese to develop a small set of rules, expressed in a tree-based pattern-action formalism, that perform local modifications of a projected analysis on the Chinese side. To avoid the slippery slope of unending language-specific rule tweaking, we strictly constrained 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$\}$).

For example, one such rule deals with noun modification:

- If $n_1, \ldots, n_k$ 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 n_k to its place with $n_1, \ldots, n_{k-1}$ as dependents.

Another deals with aspectual markers for verbs:

- If $v_1, \ldots, v_k$, a sequence of Chinese words aligned with English verbs, is followed by a, an aspect marker, make a into a modifier of the last verb v_k.

The most involved transformation places a linguistic constraint on the Chinese functional word *de*, which may be translated as *that* (the head of a relative clause), as the preposition *of*, or as *'s* (a marker for possessives). This common Chinese functional word is almost always either unaligned or multiply aligned to an English word.

- If c_i is the Chinese word that appeared immediately to the left of *de* and c_j is the Chinese word that appeared immediately to the right of it, then find the lowest ancestors c_p and c_q for c_i and c_j, respectively, such that $R(c_p, c_q)$ exists; remove that relationship; and replace it with $R(de, c_p)$ and $R(c_q, de)$.

The latter two changes may seem unrelated, but they both take advantage of the fact that Chinese violates the head-initial rule in its nominal system, where noun phrases are uniformly head-final. More generally, 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 (a few days) and express the solution compactly (< 20 rules). The complete set of rules can be found in (Hwa et al., 2002).

3.4 A New Experiment

Because our error analysis and subsequent algorithm refinements made use of our original Chinese-English data set, we created a new test set based on 88 new Chinese sentences from the Penn Chinese Treebank, already manually translated into English as part of the NIST MT evaluation preview.[5] These sentences averaged 19.0 words in length.

As described above, parses on the English side were created semi-automatically, and word alignments were acquired manually. However, in order to reduce our reliance on linguistically sophisticated human annotators for Chinese syntax, we adopted an alternative strategy for obtaining the gold standard: we automatically converted the Treebank's constituency parses of the Chinese sentences into syntactic dependency representations, using an algorithm similar to the one described in Section 2 of the paper by Xia and Palmer (2001).[6]

The recall and precision figures for the new experiment are summarized in Table 2. The first row of the table shows the results comparing the output of the Direct Projection Algorithm with the gold standard. As we have already seen previously, the quality of these trees is not very good. The second row of the table shows that after applying the single transformation based on the head-initial assumption, precision and recall both improve significantly: using the F-measure to combine precision and recall into a single figure of merit (Van Rijsbergen, 1979), the increase from 38.1% to 59.4% represents a 55.9% relative improvement. The third row of the table shows that by applying the small set of tree modification rules after direct projection (one of which is default assignment of the head-initial analysis to multi-word phrases when no other rule applies), we obtain an even larger improvement, the 67.3% F-measure representing a 76.6% relative gain over baseline performance.

Method	Precision	Recall	F-measure
Direct	34.5	42.5	38.1
Head-initial	59.4	59.4	59.4
Rules	68.0	66.6	67.3

Table 2: Performance on Chinese analyses (%)

4 Conclusions and Future Work

To what extent is the DCA a valid assumption? Our experiments confirm the linguistic intuition, indicating that one cannot safely assume a direct mapping between the syntactic dependencies of one language and the syntactic dependencies of another.

How useful is the DCA? The experimental results show that even the simplistic DCA can be useful when operating in conjunction with small quantities of systematic linguistic knowledge. 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. In the near future we will address the remaining errors, which also seem to be amenable to a uniform linguistic treatment: in large part they involve differences in category expression (nominal expressions translated as verbs or vice versa) and we believe that we can use context to effect the correct category transformations. We will also explore correction of errors via statistical learning techniques.

The implication of this work for statistical translation modeling is that a little bit of knowledge can be a good thing. The approach described here strikes a balance somewhere between the endless construction-by-construction tuning of rule-based approaches, on the one

[5] See http://www.nist.gov/speech/tests/mt/. We used sentences from sections 038, 039, 067, 122, 191, 207, 249 because, according to the distributor, the translation of these sections (files with .spc suffix) have been more carefully verified.

[6] The strategy was validated by performing the same process on the original data set; the agreement rate with the human-generated dependency trees was 97.5%. This led us to be confident that Treebank constituency parses could be used automatically to create a gold standard for syntactic dependencies.

hand, and, on the other, the development of insufficiently constrained stochastic models.

We have systematically diagnosed a common assumption that has been dealt with previously on a case by case basis, but not named. Most of the models we know of — from early work at IBM to second-generation models such as that of Knight and Yamada — rectify glaring problems caused by the failure of the DCA using a range of pre- or post-processing techniques.

We have identified the source for a host of these problems and have suggested diagnostics for future cases where we might expect these problems to arise. More important, we have shown that linguistically informed strategies can be developed efficiently to improve output that is otherwise compromised by situations where the DCA does not hold.

In addition to resolving the remaining problematic cases for our projection framework, we are exploring ways to automatically create large quantities of syntactically annotated data. This will break the bottleneck in developing appropriately annotated training corpora. Currently, we are following two research directions. Our first goal is to minimize the degree of degradation in the quality of the projected trees when the input analyses and word alignments are automatically generated by a statistical parser and word alignment model. To improve the quality of the input analyses, we are adapting active learning and co-training techniques (Hwa, 2000; Sarkar, 2001) to exploit the most reliable data. We are also actively developing an alternative alignment model that makes more use of the syntactic structure (Lopez et al., 2002). Our second goal is to detect and reduce the noise in the projected trees so that they might replace the expensive human-annotated corpora as training examples for statistical parsers. We are investigating the use of filtering strategies to localize the potentially problematic parts of the projected syntactic trees.

Acknowledgments

This work has been supported, in part, by ONR MURI Contract FCPO.810548265, DARPA/ITO Cooperative Agreement N660010028910, and Mitre Contract 010418-7712. The authors would like to thank Edward Hung, Gina Levow, and Lingling Zhang for their assistance as annotators; Michael Collins for the use of his parser; Franz Josef Och for his help with GIZA++; and Lillian Lee, the students of CMSC828, and the anonymous reviewers for their comments on this paper.

References

Anne Abeille, Kathleen Bishop, Sharon Cote, and Yves Schabes. 1990. A lexicalized tree adjoining grammar for English. Technical Report MS-CIS-90-24, University of Pennsylvania.

Hiyan Alshawi, Srinivas Bangalore, and Shona Douglas. 2000. Learning dependency transduction models as collections of finite state head transducers. *Computational Linguistics*, 26(1).

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.

Peter F. Brown, Stephen A. DellaPietra, Vincent J. DellaPietra, and Robert L. Mercer. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*.

Clara Cabezas, Bonnie Dorr, and Philip Resnik. 2001. Spanish language processing at university of maryland: Building infrastructure for multilingual applications. In *Proceedings of the Second International Workshop on Spanish Language Processing and Language Technologies (SLPLT-2)*.

Eugene Charniak. 2001. Immediate-head parsing for language models. In *Proc. of the 39th Meeting of the ACL*.

Ciprian Chelba and Fredrick Jelinek. 1998. Exploiting syntactic structure for language modeling. In *Proceedings of COLING-ACL*, volume 1, pages 225–231.

Noam Chomsky. 1981. *Lectures on Government and Binding*. Foris.

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.

Bonnie J. Dorr. 1994. Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20(4):597–635.

Jason Eisner. 1997. Bilexical grammars and a cubic-time probabilistic parser. In *Proceedings of the International Workshop on Parsing Technologies.*

Chung-Hye Han, Benoi Lavoi, Martha Palmer, Owen Rambow, Richard Kittredge, Tanya Korelsky, Nari Kim, and Myunghee Kim. 2000. Handling structural divergences and recovering dropped arguments in a Korean/English machine translation system. In *Proceedings of the AMTA*.

Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. 2002. Evaluating translational correspondence using annotation projection. Technical report, University of Maryland.

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.

Benoit Lavoie, Michael White, and Tanya Korelsky. 2001. Including Lexico-Structural Transfer Rules from Parsed Bi-texts. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics – DDMT Workshop*, Toulouse, France.

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.

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.

I. Dan Melamed. 1998. Annotation style guide for the blinker project. Technical Report IRCS 98-06, University of Pennsylvania.

Arul Menezes and Stephen D. Richardson. 2001. A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics – DDMT Workshop*, Toulouse, France.

Anoop Sarkar. 2001. Applying co-training methods to statistical parsing. In *Proc. of NAACL*, June.

Stuart Shieber. 1994. Restricting the weak-generative capacity of synchronous tree-adjoining grammars. *Computational Intelligence*, 10(4):371–385, November.

C. J. Van Rijsbergen. 1979. *Information Retrieval.* Butterworth.

Dekai Wu. 1995. Stochastic inversion transduction grammars, with application to segmentation, bracketing, and alignment of parallel corpora. In *Proc. of the 14th Intl. Joint Conf. on Artificial Intelligence*, pages 1328–1335, Aug.

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.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proc. of the Conference of the Association for Computational Linguistics*, pages 523–529.

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.