

Structure Alignment Using Bilingual Chunking

Wei Wang[†]

Beijing University of Posts and Telecomm, 181#, Beijing, 100876, P.R.C.,
y9772209@bupt.edu.cn

Jin-Xia Huang

Microsoft Research, Asia Beijing, 100080, P.R.C.
i-jxh@microsoft.com

Ming Zhou

Microsoft Research, Asia Beijing, 100080, P.R.C

mingzhou@microsoft.com

Chang-Ning Huang

Microsoft Research, Asia Beijing, 100080, P.R.C

cnhuang@microsoft.com

Abstract

A new statistical method called “bilingual chunking” for structure alignment is proposed. Different with the existing approaches which align hierarchical structures like sub-trees, our method conducts alignment on chunks. The alignment is finished through a simultaneous bilingual chunking algorithm. Using the constrains of chunk correspondence between source language (SL)¹ and target language (TL), our algorithm can dramatically reduce search space, support time synchronous DP algorithm, and lead to highly consistent chunking. Furthermore, by unifying the POS tagging and chunking in the search process, our algorithm alleviates effectively the influence of POS tagging deficiency to the chunking result.

The experimental results with English-Chinese structure alignment show that our model can produce 90% in precision for chunking, and 87% in precision for chunk alignment.

Introduction

We address here the problem of structure alignment, which accepts as input a sentence pair,

[†] This work was done while the author was visiting Microsoft Research Asia

¹ In this paper, we take English-Chinese parallel text as example; it is relatively easy, however, to be extended to other language pairs.

and produces as output the parsed structures of both sides with correspondences between them.

The structure alignment can be used to support machine translation and cross language information retrieval by providing extended phrase translation lexicon and translation templates.

The popular methods for structure alignment try to align hierarchical structures like sub-trees with parsing technology. However, the alignment accuracy cannot be guaranteed since no parser can handle all authentic sentences very well. Furthermore, the strategies which were usually used for structure alignment suffer from serious shortcomings. For instance, *parse-to-parse matching* which regards parsing and alignment as separate and successive procedures suffers from the inconsistency between grammars of different languages. *Bilingual parsing* which looks upon parsing and alignment as a simultaneous procedure needs an extra ‘bilingual grammar’. It is, however, difficult to write a complex ‘bilingual grammar’.

In this paper, a new statistical method called “bilingual chunking” for structure alignment is proposed. Different with the existing approaches which align hierarchical structures like sub-trees, our method conducts alignment on chunks. The alignment is finished through a simultaneous bilingual chunking algorithm. Using the constrains of chunk correspondence between source language (SL) and target language (TL), our algorithm can dramatically reduce search space, support time synchronous DP algorithm,

and lead to highly consistent chunking. Furthermore, by unifying the POS tagging and chunking in the search process, our algorithm alleviates effectively the influence of POS tagging deficiency to the chunking result.

The experimental results with English- Chinese structure alignment show that our model can produce 90% in precision for chunking, and 87% in precision for chunk alignment.

1 Related Works

Most of the previous works conduct structure alignment with complex, hierarchical structures, such as phrase structures (e.g., Kaji, Kida & Morimoto, 1992), or dependency structures (e.g., Matsumoto *et al.* 1993; Grishman, 1994; Meyers, Yanharber & Grishman 1996; Watanabe, Kurohashi & Aramaki 2000). However, the mismatching between complex structures across languages and the poor parsing accuracy of the parser will hinder structure alignment algorithms from working out high accuracy results.

A straightforward strategy for structure alignment is *parse-to-parse matching*, which regards the parsing and alignment as two separate and successive procedures. First, parsing is conducted on each language, respectively. Then the correspondent structures in different languages are aligned (e.g., Kaji, Kida & Morimoto 1992; Matsumoto *et al.* 1993; Grishman 1994; Meyers, Yanharber & Grishman 1996; Watanabe, Kurohashi & Aramaki 2000). Unfortunately, automatic *parse-to-parse matching* has some weaknesses as described in Wu (2000). For example, grammar inconsistency exists across languages; and it is hard to handle multiple alignment choices.

To deal with the difficulties in *parse-to-parse matching*, Wu (1997) utilizes inversion transduction grammar (ITG) for *bilingual parsing*. *Bilingual parsing* approach looks upon the parsing and alignment as a single procedure which simultaneously encodes both the parsing and transferring information. It is, however, difficult to write a broad coverage ‘bilingual grammar’ for *bilingual parsing*.

2 Structure Alignment Using Bilingual Chunking

2.1 Principle

The chunks, which we will use, are extracted from the Treebank. When converting a tree to the chunk sequence, the chunk types are based on the syntactic category part of the bracket label. Roughly, a chunk contains everything to the left of and including the syntactic head of the constituent of the same name. Besides the head, a chunk also contains pre-modifiers, but no post-modifiers or arguments (Erik. 2000).

Using chunk as the alignment structure, we can get around the problems such as PP attachment, structure mismatching across languages. Therefore, we can get high chunking accuracy.

Using bilingual chunking, we can get both high chunking accuracy and high chunk alignment accuracy by making the SL chunking process and the TL chunking process constrain and improve each other.

Our ‘bilingual chunking’ model for structure alignment comprises three integrated components: *chunking models* of both languages, and the *crossing constraint*; it uses chunk as the structure. (See **Fig. 1**)

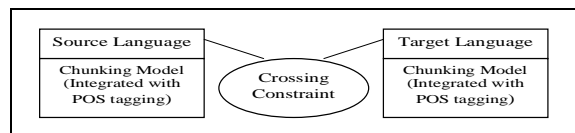


Fig. 1 Three components of our model

The *crossing constraint* requests a chunk in one language only correspond to at most one chunk in the other language. For instance, in **Fig. 2** (the dashed lines represent the word alignments; the brackets indicate the chunk boundaries), the phrase “*the first man*” is a monolingual chunk, it, however, should be divided into “*the first*” and “*man*” to satisfy the *crossing constraint*. By

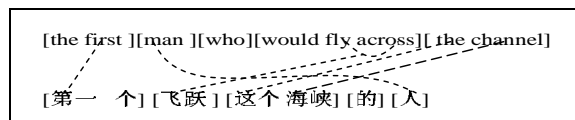


Fig. 2 the crossing constraint

using *crossing constraint*, the illegal chunk candidates can be removed in the chunking process.

The chunking models for both languages work successively under the *crossing constraint*. Usually, chunking involves two steps: (1) POS tagging, and (2) chunking. To alleviate effectively the influence of POS tagging deficiency to the chunking result, we integrate the two steps with a unified model for optimal solution. This integration strategy has been proven to be effective for base NP identification (Xun, Huang & Zhou, 2001).

Consequently, our model works in three successive steps: (1) word alignment between SL and TL sentences; (2) source language chunking; (3) target language chunking. Both (2) and (3) should work under the supervision of *crossing constraints*.

2.2 The Crossing Constraint

According to (Wu, 1997), *crossing constraint* can be defined in the following.

For non-recursive phrases: Suppose two words w_1 and w_2 in language-1 correspond to two words v_1 and v_2 in language-2, respectively, and w_1 and w_2 belong to the same phrase of language-1. Then v_1 and v_2 must also belong to the same phrase of language-2.

We can benefit from applying *crossing constraint* in the following three aspects:

- *Consistent chunking in the view of alignment.* For example, in **Fig. 2**, “*the first man*” should be divided into “*the first*” and “*man*” for the consistency with the Chinese chunks “*第一个*” and “*人*”, respectively.
- *Searching space reduction.* The chunking space is reduced by ruling out those illegal fragments like “*the first man*”; and the alignment space is reduced by confining those legal fragments like “*the first*” only to correspond to the Chinese fragments “*第一*” or “*第一个*” based on word alignment anchors.
- *Time synchronous algorithms for structure alignment.* Time synchronous algorithms cannot be used due to word permutation problem before. While under the *crossing constraint*, these

algorithms (for example, dynamic programming) can be used for both chunking and alignment.

2.3 Mathematical Formulation

Given an English sentence $e = w_e^1 w_e^2 \dots, w_e^l$, its POS tag sequence is denoted by $T_e = t_e^1 t_e^2 \dots, t_e^l$, where l is the sentence length. A sequence of chunks can be represented as:

$$B_e = [t_e^1 t_e^2], \dots, [t_e^i t_e^{i+1} t_e^{i+2}], \dots, [t_e^l] = n_e^1, \dots, n_e^{l'}$$

Where, n_e^i denotes the i^{th} chunk type of e , and l' is the number of chunks in e .

Similarly, for a Chinese sentence c

$$c = w_c^1 w_c^2 \dots, w_c^m$$

$$T_c = t_c^1 t_c^2 \dots, t_c^m$$

$$B_c = [t_c^1 t_c^2], \dots, [t_c^i t_c^{i+1} t_c^{i+2}], \dots, [t_c^m] = n_c^1, \dots, n_c^{m'}$$

Where, m denotes the word number of c , m' is the number of Chinese chunks in c .

Let bmi denote the i^{th} positional tag, bmi can be *begin* of a chunk, *inside* a chunk, or *outside* any chunk.

The most probable result is expressed as $\langle B_e^*, B_c^*, A^* \rangle = \arg \max_{B_e, B_c, A} p(B_e, B_c, A, c | e, a)$ (1)

Where, A is the alignment between B_e and B_c . a refers to the *crossing constraint*. Equation (1) can be further derived into

$$\begin{aligned} & \langle B_e^*, B_c^*, A^* \rangle \\ &= \arg \max_{B_e, B_c, A} \sum_{T_e} \sum_{T_c} p(B_e, B_c, A, T_e, T_c, c | e, a) \\ &\approx \arg \max_{B_e, B_c, A, T_e, T_c} p(B_e, B_c, A, T_e, T_c, c | e, a) \end{aligned} \quad (2)$$

Using Bayes' rule, equation (2) will be

$$\begin{aligned} & \langle B_e^*, B_c^*, A^* \rangle = \arg \max_{B_e, B_c, T_e, T_c} \{ \\ & p(T_e | e, a) p(B_e | T_e, e, a) \\ & \times p(T_c | B_e, T_e, e, a) \times p(c | T_c, B_e, T_e, e, a) \\ & \times p(B_c | c, T_c, B_e, T_e, e, a) \times p(A | B_c, c, T_c, B_e, T_e, e, a) \} \end{aligned} \quad (3)$$

In this formula, $p(T_e | e, a)$ aims to determine the best POS tag sequence for e . $p(B_e | T_e, e, a)$ aims to determine the best chunk sequence from them. $p(c | T_c, B_e, T_e, e, a)$ aims to decide the best POS tag sequence for c based on the English POS sequence. $p(B_c | c, T_c, B_e, T_e, e, a)$ aims to decide the best Chinese chunking result based on the Chinese POS sequence and the English chunk sequence.

Note that $p(A | B_c, c, T_c, B_e, T_e, e, a) = 1$

In practice, in order to reduce the search space, only N-best results of each step are retained.

Determining the N-Best English POS Sequences

The HMM based POS tagging model (Kupiec 1992) with the trigram assumption is used to provide possible POS candidates for each word in terms of the N-best lattice.

Determining the N-best English Chunking Result

This step is to find the best chunk sequence based on the N-best POS lattice by decomposing the chunking model into two sub-models (1) inter-chunk model; (2) intra- chunk model.

From equation (3), based on Bayes' rule, then

$$B_e(M - best) = \arg \max_{\substack{B_e \\ T_e \in T_e(N-best)}} (p(B_e | a) \times p(T_e | B_e) \times p(e | B_e, T_e, a)) \quad (4)$$

Based on the trigram assumption, the first part can be written as,

$$p(B_e | a) = \prod_{i=1}^{l'} p(n_e^i | n_e^{i-2}, n_e^{i-1}, a) \quad (5)$$

Here, the *crossing constraint* a will remove those illegal candidates.

The second part can be further derived based on two assumptions: (1) bigram for the English POS transition inside a chunk; (2) the first POS tag of a chunk only depends on the previous two tags. Thus

$$p(T_e | B_e) = \prod_{i=1}^{l'} (p(t_e^{i,1} | t_e^{i-2}, t_e^{i-1}) \prod_{j=1}^{x_i} p(t_e^{i,j} | t_e^{i,j-1}, n_e^i)) \quad (6)$$

Where, x_i is the number of words that the i^{th} English chunk contains. And $t_e^{i,-2}, t_e^{i,-1}$ refer to the two tags before $t_e^{i,1}$.

The third part can be derived based on the assumption that an English word w_e^i only depends on its POS tag (t_e^i), chunk type (n_e^i) it belongs to and its positional information (bm_e^i) in the chunk, thus

$$p(e | B_e, T_e, a) = \prod_{i=1}^l p(w_e^i | t_e^i, bm_e^i, n_e^i) \quad (7)$$

i' is the index of the chunk the word belongs to.

Finally, from (4)(5)(6)(7), we arrive

$$B_e(N - best) = \arg \max_{\substack{B_e \\ T_e \in T_e(N-best)}} \prod_{i=1}^{l'} \left\{ \frac{p^\beta(n_e^i | n_e^{i-2}, n_e^{i-1}, a) p(t_e^{i,1} | t_e^{i-2}, t_e^{i-1})}{\text{inter-chunk prob.}} \times \prod_{j=1}^{x_i} \frac{(p(w_e^{i,j} | t_e^{i,j}, bm_e^{i,j}, n_e^i) p(t_e^{i,j} | t_e^{i,j-1}, n_e^i))}{\text{intra-chunk prob.}} \right\} \quad (8)$$

Where β is a normalization coefficient, and its value is 0.5 in our experiment.

Deciding the Chinese N-best POS Sequences

The N-best Chinese POS sequences are obtained by considering four factors: (1) tag transition probability; (2) tag translation probability; (3) lexical generation probability; (4) lexicon translation probability.

From Equation (3), we can derive

$$p(T_c | T_e, a) = \prod_{i=1}^m \left(\underbrace{p(t_c^i | t_c^{i-2}, t_c^{i-1})}_{\text{POS tag transition prob.}} \times \underbrace{\left\{ \frac{p(t_c^i | t_e^i, (i,j) \in \text{conn})}{\mathbb{1}_{\text{if } (i,j) \notin \text{conn}}} \right\}}_{\text{POS tag translatio n prob.}} \right) \quad (9)$$

Where, *conn* is the word alignment result. And

$$p(c | T_e, e, a) = \prod_{i=1}^m \left(\underbrace{p(w_c^i | t_c^i)}_{\text{lex generation prob.}} \times \underbrace{\left\{ \frac{p(w_c^i | w_e^i, (i,j) \in \text{conn})}{\mathbb{1}_{\text{if } (i,j) \notin \text{conn}}} \right\}}_{\text{lex translatio n prob.}} \right) \quad (10)$$

We assume the word translation probability is 1 since we are using the word alignment result. Comparing with a typical HMM based tagger, our model also utilizes the POS tag information in the other language.

Obtaining the Best Chinese Chunking Result

Similar to the English chunking model, the Chinese chunking model also includes (1) inter-chunk model; (2) intra-chunk model. They are simplified, however, because of limited training data.

Using the derivation similar to equation (4)–(8), we can get (11) form equation (3) with the assumptions that (1) $p(B_c | c, T_c, B_e, T_e, e, a)$ depends

only on T_c, c and a ; (2) bigram for chunk type transition; (3) bigram for tag transition inside a chunk; (4) trigram for the POS tag transition between chunks, we get

$$B_c^* = \arg \max_{\theta, T_c} \prod_{i=1}^m \underbrace{[p(n_c^i | n_c^{i-1}) p(t_c^{i,1} | t_c^{i-1}, t_c^{i-2})]}_{\text{inter-chunk prob}} \times \prod_{j=1}^i \underbrace{p(t_c^{i,j} | t_c^{i,j-1}, n_c^i)}_{\text{intra-chunk prob}} \quad (11)$$

i ' is the word number of the i^{th} Chinese phrase.

2.4 Model Estimation

We use three kinds of resources for training and testing: **a)** The WSJ part of the Penn Treebank II corpus (Marcus, Santorini & Marcinkiewicz 1993). Sections 00-19 are used as the training data, and sections 20-24 as the test data. **b)** The HIT Treebank², containing 2000 sentences. **c)** The HIT bilingual corpus³, containing 20,000 sentence-pairs (in general domain) annotated with POS and word alignment information. We used 19,000 sentence-pairs for training and 1,000 for testing. These 1000 sentence-pairs are manually chunked and aligned.

From the Penn Treebank, English chunks were extracted with the conversion tool (<http://lcg-www.uia.ac.be/conll2000/chunking>). From the HIT Treebank, Chinese chunks were extracted with a conversion tool implemented by ourselves. We can obtain an English chunk bank and a Chinese chunk bank.

With the chunk dataset obtained above, the parameters were estimated with Maximum Likelihood Estimation.

The POS tag translation probability in equation (9) was estimated from **c)**.

The English part-of-speech tag set is the same with Penn Treebank. And the Chinese tag set is the same with HIT Treebank.

13 chunk types were used for English, which are the same with (Erik et al, 2000). 7 chunk types were used for Chinese, including BDP (adverb phrase), BNP (noun phrase), BAP (adjective

phrase), BVP (verb phrase), BMP (quantifier phrase), BPP (prepositional phrase) and O (words outside any other chunks).

3 Experimental Results

We conducted experiments to evaluate (1) the overall accuracy; (2) the comparison with isolated strategy; (3) the comparison with a score-function approach.

The word aligner developed by Wang *et al.* (2001) was used to provide word alignment anchors. The 1000 sentence-pairs described in section 2.4 were used as evaluation standard set.

The result is evaluated in terms of chunking precision and recall, as well as alignment precision and recall, as defined in the following:

$$\begin{aligned} \text{Chunking Pre.} &= \frac{\# \text{ chunks correctly identified}}{\# \text{ chunks identified}} \\ \text{Chunking Rec.} &= \frac{\# \text{ chunks correctly identified}}{\# \text{ chunks should be identified}} \\ \text{Alignment Pre.} &= \frac{\# \text{ Eng. chunks correctly aligned}}{\# \text{ Eng. chunks aligned}} \\ \text{Alignment Rec.} &= \frac{\# \text{ Eng. chunks correctly aligned}}{\# \text{ Eng. chunks should be aligned}} \end{aligned}$$

3.1 Overall Accuracy

As described in section 2.3, in each step, N-best candidates were selected. In our experiment, N

Table 1: Overall accuracy

	English Chunking		Chinese Chunking		Alignment	
	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
N=1	90.34	90.67	88.41	87.05	85.31	81.07
N=2	92.34	92.93	89.52	88.80	86.54	82.69
N=3	93.21	94.16	89.90	89.58	86.96	83.58
N=4	93.48	94.94	89.93	90.11	87.05	84.16
N=5	92.91	94.43	89.41	89.77	86.69	83.89
N=6	92.70	94.20	89.29	89.72	86.57	83.79
N=7	92.31	93.88	88.89	89.46	86.17	83.51

was set from 1 to 7.

Table 1 shows the results with different N. When N=4, we get the best results, we got 93.48% for English chunking, 89.93% for Chinese chunking, and 87.05% for alignment.

² <http://mtlab.hit.edu.cn/download/4.TXT>

³ Created by Harbin Institute of Technology.

Table 2 shows the results of individual Chinese chunk types. The second column is the percentage that each type occupies among all the Chinese chunks. Table 3 shows the results of individual English chunk types. The last column shows the alignment precision of each English

Table 2 : accuracy of Chinese chunk types

Chunk Type	% in corpus	Pre. %	Rec. %
BNP	34.60	89.25	92.49
BVP	23.50	84.66	87.03
BPP	4.85	88.54	87.04
BDP	5.99	90.13	91.78
BAP	2.86	83.49	84.69
BMP	1.30	73.45	87.37
O	26.89	98.02	90.65

chunk type.

We can see from table 2 and 3 that the precision and recall for chunks of NP, BNP, ADVP, DP, and O are around 90% for both Chinese and English. This reflects that the compositional rules of these chunk types are very regular.

Table 3 : accuracy of English chunk types

Chunk Type	% in corpus	Chunking		Alignment
		Pre. %	Rec. %	Pre. %
NP	39.34	93.84	95.83	89.08
VP	20.02	90.67	90.12	80.66
PP	11.48	92.32	95.78	75.64
ADVP	4.02	92.67	92.98	86.11
SBAR	1.28	92.08	97.89	86.27
ADJP	2.49	86.00	92.97	83.43
PRT	1.08	87.34	86.25	62.96
INTJ	0.05	97.06	94.26	100.00
O	19.81	97.77	98.51	91.61

3.2 Chunking Ability Evaluation: Comparison with Isolated Strategy

We now compare with the *isolated strategy*, which separately conduct chunks for English and Chinese.

In *isolated strategy*, we carry out the English and Chinese chunking separately, we call this experiment **M**.

We next add the *crossing constraint* to **M**. In other words, chunk each language under the *crossing constraint*, without considering the chunking procedure of the correspondent language. We call this experiment **M+C**.

Both **M** and **M+C** are compared with our integrated mode, which we call **I**.

Table 4 chunking accuracies of different approaches

	English Chunking Accuracy		Chinese Chunking Accuracy	
	Pre. %	Rec. %	Pre. %	Rec. %
M	92.52	90.81	72.30	81.60
M+C	92.84	92.68	79.88	83.61
I	93.48	94.94	89.93	90.11

Table 4 indicates the contribution of the *crossing constraint* and our integrated strategy. Comparing **M+C** with **M**, we see that the accuracies (pre. & rec.) of both languages rise. Comparing **I** with **M+C**, the accuracies rise again.

Table 5 searching space of different approaches

	English (#chunk candidate)	Chinese (#chunk candidate)
M	59790	57043
M+C	46937	14746
I	46937	14746

In table 5, please note that the searching spaces of **M+C** and **I** are the same. This is because they all adopt the *crossing constraint*. Comparing both **I** and **M+C** with **M**, we see that the searching space is reduced 21% ((59790-46937)/59790) for English and 71% ((57043-14746)/57043) for Chinese and 47% ((59790+57043-46937-14746) / (59790+57043)) for all.

3.3 Alignment Evaluation: Comparing with Score Function Approach

The score-function approach is usually used to select the best target language correspondence for a source language fragment. Here, we call it **SF**.

First, we parse the English side under the *crossing constraint* (as the **M+C** case in section 3.2). And then use a score function to find the target correspondence for each English chunk.

The score function is:

$$SF = p(m | l) p(\Delta k | m, l) p(\Delta j | m, l)$$

m and l are the lengths of the English chunk and its correspondent Chinese chunk respectively. Δk is the difference in number of content words between these two chunks, Δj is the difference of functional words. This function achieves the best performance among several

lexicalized score functions in (Wang, *et al.*, 2001). The alignment result is shown in table 6.

Table 6 : finding target correspondence

SF		I (ntegrated)	
Pre. %	Rec. %	Pre. %	Rec. %
68.33	66.12	87.05	84.16

The comparison between **SF** and **I** indicates that our integrated model obviously outperforms the score function approach in the aspect of finding the target alignment for source language chunks.

Conclusion

A new statistical method called “bilingual chunking” for structure alignment is proposed. Different with the existing approaches which align hierarchical structures like sub-trees, our method conducts alignment on chunks. The alignment is finished through a simultaneous bilingual chunking algorithm. Using the constrains of chunk correspondence between source language (SL) and target language(TL), our algorithm can dramatically reduce search space, support time synchronous DP algorithm , and lead to highly consistent chunking. Furthermore, by unifying the POS tagging and chunking in the search process, our algorithm alleviates effectively the influence of POS tagging deficiency to the chunking result.

The experimental results with English-Chinese structure alignment show that our model can produce 90% in precision for chunking, and 87% in precision for chunk alignment. Compared with the isolated strategy, our method achieves much higher precision and recall for bilingual chunking. Compared with the score function approach, our method got much higher precision and recall for chunk alignment.

In the future, we will conduct further research such as the inner-phrase translation modeling, or transferring grammar introduction, bilingual pattern learning, etc, based on the results of our method.

References

Erik F. Tjong Kim Sang and Sabine Buchholz (2000) *Introduction to the CoNLL-2000 Shared Task:*

Chunking. CoNL-2000 and LLL-2000. Lisbon, Portugal, pp. 127-132.

Grishman R. (1994) *Iterative Alignment of Syntactic Structures for a Bilingual Corpus*. WVLC-94, pp. 57-68.

Huang, J. and Choi, K. (2000) *Chinese-Korean Word Alignment Based on Linguistic Comparison*. ACL-2001.

Kaji, H., Kida, Y., and Morimoto, Y. (1992) *Learning Translation Templates from Bilingual Texts*. COLING-92, pp. 672-678.

Matsumoto, Y., Ishimoto, H., and Utsuro, T. (1993) *Structural Matching of Parallel Texts*, ACL-93, pp. 23-30.

Kupiec J. (1992) *Robust Part-of-speech tagging using a hidden Markov model*. Computer Speech and Language 6.

Meyers, A., Yanharber, R., and Grishman, R. (1996) *Alignment of Shared Forests for Bilingual Corpora*. Colings-96, pp. 460-465.

Wang Wei, Huang Jin-Xia, Zhou Ming and Huang Chang-Ning (2001) *Finding Target Language Correspondence for Lexical EBMT system*. NLP RS-2001.

Wang Y. and Waibel A. (1998) *Modeling with Structures in Statistical Machine Translation*. COLING-ACL 1998.

Watanabe H., Kurohashi S., Aramaki E. (2000) *Finding Structural Correspondences from Bilingual Parsed Corpus for Corpus-based Translation*. COIING-2000.

Wu Dekai (2000) *Alignment*. Handbook of Natrual Language Processing, Robet Dale, Hermann Moisl, and Harold Somers ed, Marcel Dekker, Inc. pp. 415-458,

Wu, Dekai (1997) *Stochastic inversion transduction grammars and bilingual parsing of parallel corpora*. Computational Linguistics 23/3, pp. 377-404.

Xun, E., Huang, C., and Zhou, M. (2001) *A Unified Statistical Model for the Identification of English BaseNP*. ACL-2001.