

# A Robust Cross-Style Bilingual Sentences Alignment Model

Tz-Liang Kueng      Keh-Yih Su

Behavior Design Corporation  
2F, No.5, Industry E. Rd. IV,  
Science-Based Industrial Park,  
Hsinchu, Taiwan 30077, R.O.C.  
{cavs, kysu}@bdc.com.tw

## Abstract

Most current sentence alignment approaches adopt *sentence length* and *cognate* as the alignment features; and they are mostly trained and tested in the documents with the same style. Since the length distribution, alignment-type distribution (used by length-based approaches) and cognate frequency vary significantly across texts with different *styles*, the length-based approaches fail to achieve similar performance when tested in corpora of different styles. The experiments show that the performance in *F*-measure could drop from 98.2% to 85.6% when a length-based approach is trained by a technical manual and then tested on a general magazine.

Since a large percentage of content words in the source text would be translated into the corresponding translation duals to preserve the meaning in the target text, transfer lexicons are usually regarded as more reliable cues for aligning sentences when the alignment task is performed by human. To enhance the robustness, a *robust* statistical model based on both transfer lexicons and sentence lengths are proposed in this paper. After integrating the transfer lexicons into the model, a 60% *F*-measure error reduction (from 14.4% to 5.8%) is observed.

## 1 Introduction

Since the bilingual corpus is a valuable resource for training statistical language models [Dagon, 91; Su et al., 95; Su and Chang, 99] and sentence alignment is the first step for most such tasks, many alignment approaches have been proposed in the literature [Brown, 91; Gale and Church, 93; Wu, 94; Vogel et al., 96; Och and Ney, 2000]. Most of those reported approaches use the sentence length as the main feature to perform the alignment task. For example, Brown et al. (91) used the feature

of number-of-words for alignment, and [Gale and Church,93] claimed that better performance can be achieved (5.8% error rate for English-French corpus) if the number-of-characters is adopted instead. As cognates are reliable cues for language pairs derived from the same family, Church (93) also attacked this problem by considering cognates additionally. Because most of those reported work are performed on those Indo-European language-pairs, for testing the performance on non-Indo-European languages, Wu (94) had tried both length and cognate features on the Hong Kong Hansard English-Chinese corpus, and 7.9% error rate has been reported. Besides, sentence alignment can also be indirectly achieved via more complicated word corresponding models [Brown et al., 93; Vogel et al., 96; Och and Ney, 2000]. Since those word corresponding models, which also achieve similar performance, are more complicated and run relatively slow, they seems to be over-killed for the task of aligning sentences and will not be discussed in this paper.

Although length-based approaches above mentioned are simple and can achieve good performance, they are usually trained and tested in the text with the same style. Therefore, they are style-dependent approaches. Since performing supervised-training for each style is not feasible in many applications, it would be interesting to know whether those length-based approaches can still achieve the similar performance if they are tested in the text with different styles other than the training corpora. An experiment was thus conducted to train the parameters with a machinery technical manual; the performance is then tested on a general magazine (for introducing Taiwan to foreign visi-

tors). It shows that the testing set performance of the length-based model (with cognates considered) would drop from 98.2% (tested in the same technical domain) to 85.6% (tested in the new general magazine) in  $F$ -measure. After investigating those errors, it has been found that the length distribution and alignment-type distribution (used by those length-based approaches) vary significantly across the texts of different styles (as would be shown in Tables 5.2 and 5.3), and the cognate-frequency<sup>1</sup> drops greatly from the technical manual to a general magazine in non-Indo-European languages (as would be shown in Table 5.3).

On the other hand, sentence length is seldom used by a human to align bilingual sentences. They usually do not align bilingual sentences by counting the number of characters (or words) in the sentence pairs. Instead, since a large percentage of content words in the source text would be translated into their translation-duals to preserve the meaning in the target text, transfer-lexicons are usually used for aligning sentences when the alignment task is performed by human. To enhance the robustness across different styles, transfer-lexicons are thus integrated into the traditional sentence-length based model in the proposed robust statistical model described below. After integrating transfer-lexicons into the model, a 60%  $F$ -measure error reduction (from 14.4% to 5.8%) has been observed, which corresponds to improving the cross-style performance from 85.6% to 94.2% in  $F$ -measure.

The details of the proposed robust model, the associated features extracted from the bilingual corpora, and the probabilistic scoring function will be given in Section 2. In Section 3, we briefly mention some implementation issues. The associated performance evaluation is given in Section 4, and Section 5 would address error analysis and discusses the limitation of the proposed statistical model. Finally, the concluding remarks are given in Section 6.

<sup>1</sup>Here "Cognate" mainly refers to those English proper nouns (such as those company names of IBM, HP; or the technical terms such as IEEE-1394, etc.) that appear in the Chinese text. As they are most likely to be directly copied from the English sentence into the corresponding Chinese one, they are reliable cues.

## 2 Statistical Sentence Alignment Model

Since an English-Chinese bilingual corpus will be adopted in our experiments, we will denote the source text with  $m$  sentences as  $ES_1^m$ , and its corresponding target text, with  $n$  sentences, as  $CS_1^n$ . Let  $M_i = \{type_{i,1}, \dots, type_{i,N_i}\}$  denote the  $i$ -th possible alignment-candidate, consisting of  $N_i$  *Alignment-Passages* of  $type_{i,j}$ ,  $j = 1, \dots, N_i$ ; where  $type_{i,j}$  is the matching type (e.g., 1-1, 0-1, 1-0, etc.) of the  $j$ -th *Alignment-Passage* in the  $i$ -th alignment-candidate, and  $N_i$  denotes the number of the total *Alignment-Passages* in the  $i$ -th alignment-candidate. Then the statistical alignment model is to find the Bayesian estimate  $M^*$  among all possible alignment candidates, shown in the following equation

$$M^* = \arg \max_{M_i} P(M_i | ES_1^m, CS_1^n). \quad (2.1)$$

According to the Bayesian rule, the maximization problem in (2.1) is equivalent to solving the following maximization equation

$$\begin{aligned} M^* &= \arg \max_{M_i} P(ES_1^m, CS_1^n | M_i) P(M_i) \\ &= \arg \max_{M_i} \{P(\textit{Aligned-Pair}_{i,1}^{i,N_i} | type_{i,1}^{i,N_i}) P(type_{i,1}^{i,N_i})\} \\ &= \arg \max_{M_i} \prod_{j=1}^{N_i} \{P(\textit{Aligned-Pair}_{i,j} | \textit{Aligned-Pair}_{i,1}^{i,j-1}, type_{i,1}^{i,j}) \times \\ &\quad P(type_{i,j} | type_{i,1}^{i,j-1})\}, \end{aligned} \quad (2.2)$$

where  $\textit{Aligned-Pair}_{i,j}$ ,  $j = 1, \dots, N_i$ , denotes the  $j$ -th aligned English-Chinese bilingual sentence groups pair in the  $i$ -th alignment candidate.

Assume that

$$\begin{aligned} P(\textit{Aligned-Pair}_{i,j} | \textit{Aligned-Pair}_{i,1}^{i,j-1}, type_{i,1}^{i,j}) \\ \approx P(\textit{Aligned-Pair}_{i,j} | type_{i,j}), \end{aligned} \quad (2.3)$$

and different  $type_{i,j}$  in the  $i$ -th alignment candidate are statistically independent<sup>2</sup>, then the above maximization problem can be approached by searching for

$$\hat{M} \equiv \arg \max_{M_i} \prod_{j=1}^{N_i} \{P(\textit{Aligned-Pair}_{i,j} | type_{i,j}) P(type_{i,j})\}, \quad (2.4)$$

where  $\hat{M}$  denotes the desired candidate.

<sup>2</sup>A more reasonable one should be the first-order Markov model (i.e., Type-Bigram model); however, it will significantly increase the searching time and thus is not adopted in this paper.

## 2.1 Baseline Model

To make the above model feasible, *Aligned-Pair<sub>i,j</sub>* should be first transformed into an appropriate feature space. The baseline model will use both the length of sentence [Brown et al., 91; Gale and Church, 93] and English cognates [Wu, 94], and is shown as follows:

$$\arg \max_{M_i} \prod_{j=1}^{N_i} f(\delta_c, \delta_w | type_{i,j}) P(\delta_{cognate}) P(type_{i,j}), \quad (2.5)$$

where  $\delta_c$  and  $\delta_w$  denote the normalized differences of characters and words as explained in the following;  $\delta_c$  is defined to be  $(l_{tc} - cl_{sc}) / \sqrt{l_{sc}s_c^2}$ , where  $l_{sc}$  and  $l_{tc}$  are the character numbers of the aligned bilingual portions of source text and target text, respectively, under consideration;  $c$  denotes the proportional constant for target-character-count and  $s_c^2$  denotes the corresponding target-character-count variance per source-character. Similarly,  $\delta_w$  is defined to be  $(l_{tw} - wl_{sw}) / \sqrt{l_{sw}s_w^2}$ , where  $l_{sw}$  and  $l_{tw}$  are the word numbers of the aligned bilingual portions of source text and target text, respectively;  $w$  denotes the proportional constant for target-word-count and  $s_w^2$  denotes the corresponding target-word-count variance per source-word. Also, the random variables  $\delta_c$  and  $\delta_w$  are assumed to have bivariate normal distribution and each possesses a standard normal distribution with mean 0 and variance 1. Furthermore,  $\delta_{cognate}$  denotes (“Number of English cognates found in the given Chinese sentences” – “Number of corresponding English cognates found in the given English sentences”), and is Poisson<sup>3</sup> distributed independent of its associated matching-type; also assume that  $\delta_{cognate}$  is independent of other features (i.e., character-count and word-count).

## 2.2 Proposed Transfer Lexicon Model

Since transfer-lexicons are usually regarded as more reliable cues for aligning sentences when the alignment task is performed by human, the above baseline model is further enhanced by adding

<sup>3</sup>Since almost all those English cognates found in the given Chinese sentences can be found in the corresponding English sentences,  $\delta_{cognate}$  had better to be modeled as a Poisson distribution for a rare event (rather than Normal distribution as some papers did).

those associated transfer lexicons to it. Those translated Chinese words, which are derived from each English word (contained in given English sentences) by looking up some kinds of dictionaries, can be viewed as transfer-lexicons because they are very likely to appear in the translated Chinese sentence. However, as the distribution of various possible translations (for each English lexicon) found in our bilingual corpus is far more diversified<sup>4</sup> compared with those transfer-lexicons obtained from the dictionary, only a small number of transfer-lexicons can be matched if the exact-match is specified. Therefore, each Chinese-Lexicon obtained from the dictionary is first augmented with its associated Chinese characters, and then the augmented transfer-lexicons set are matched with the target Chinese sentence(s). Once an element of the augmented transfer-lexicons set is matched in the target Chinese sentence, it is counted as being matched. So we compute the *Normalized-Transfer-Lexicon-Matching-Measure*,  $\delta_{Transfer-Lexicons}$  which denotes [(“Number of augmented transfer-lexicons matched” – “Number of augmented transfer-lexicons unmatched”) / “Total Number of augmented transfer-lexicons sets”], and add it to the original model as another additional feature.

Assume follows normal distribution and the associated parameters are estimated from the training set, Equation (2.5) is then replaced by

$$\arg \max_{M_i} \prod_{j=1}^{N_i} \{f_1(\delta_c, \delta_w | type_{i,j}) P(\delta_{cognate}) \times f_2(\delta_{Transfer-Lexicons}) \times P(type_{i,j})\}. \quad (2.6)$$

## 3 Implementation

The best bilingual sentence alignment in those above models can be found by utilizing a dynamic programming algorithm, which is similar to the dynamic time warping algorithm used in speech recognition [Rabiner and Juang, 93]. Currently, the

<sup>4</sup>For example, the English word “number” are found to be translated into “零件”, “组件”, “部件”, “零组件”, “零部件”, “部分”, ... etc., for a specific sense in the given corpus; however, the transfer entries listed in the dictionary are “零件” and “部分” only.

### Case I (Length-Type Error)

- (E1) Compared to this, modern people have relatively better nutrition and mature faster, working women marry later, and there has been a great decrease in frequency of births, so that the number of periods in a lifetime correspondingly increases, so it is not strange that the number of people afflicted with endometriosis increases greatly.
- (C1) 相對的，現代人營養好、較早熟，職業婦女又晚婚，生育次數更是大減，一生中月經來潮次數相形增加，子宮內膜異位患者增多也不足為奇了。
- (E2) The problem is not confined to women.
- (E3) “Sperm activity also noticeably decreases in men over forty,” says Taipei Medical College urologist Chang Han-sheng.
- (C2) 不只是女性，「男性在四十歲以後，精蟲活動也會明顯減少很多」臺北醫學院泌尿科主任江漢聲說。

### Case II (Length&Lexicon-Type Error)

- (E1) Second, the United States as well as Japan have provided lucrative export markets for countries in this region.
- (E2) The U.S. was particularly generous in the postwar years, keeping its markets open to products from Asia and giving nascent industries in the region a chance to catch up.
- (C1) 其次，戰後的美國完全開放進口亞洲產品，使亞洲地區的新興國家有機會發展。

Figure 1: An illustration of length&lexical type error

maximum number of either source sentences or target sentences allowed in each alignment unit is set to be “4” (i.e., we will not consider those matching-types of “5 - 1”, “5 - 2”, “1 - 5”, etc).

Let  $\{s_1, \dots, s_m\}$  and  $\{t_1, \dots, t_n\}$  be the parallel bilingual source and target sentences, and let  $S(m, n)$  be the maximum accumulated score between  $\{s_1, \dots, s_m\}$  and  $\{t_1, \dots, t_n\}$  under the best alignment path. Then  $S(m, n)$  can be evaluated recursively with the initial condition of  $S(0, 0) = 0$  in the following way:

$$S(m, n) = \max_{0 \leq h, k \leq 4} S(m - h, n - k) + score(h, k), \quad (3.1)$$

where  $score(h, k)$  denotes the local scoring function to evaluate the local passage of matching type “ $h - k$ ”.

## 4 Performance Evaluation

In the experiments, a training set consisting of 7,331 pairs of bilingual sentences, and a testing set with 1,514 pairs of bilingual sentences are extracted from the Caterpillar User Manual which is mainly about machinery. The cross-style testing set contains 274 pairs of bilingual sentences selected from the Sinorama Magazine, which is a general magazine (for introducing Taiwan to foreign visitors) with its topics covering law, politics, education, technology, science, etc. Figure 1 is an illustration of bilingual Sinorama Magazine texts.

For comparing the performance of alignment, both precision rate ( $p$ ) and recall rate ( $r$ ), defined as follows, are measured; however, only their associated  $F$ -measure<sup>5</sup> is reported for saving space.

$$p = \frac{[\text{Number of correct alignment-passages in system output}]}{[\text{Total number of all alignment-passages generated from system output}]}, \quad (4.1)$$
$$r = \frac{[\text{Number of correct alignment-passages in system output}]}{[\text{Total number of all alignment-passages contained in benchmark corpus}]}. \quad (4.2)$$

A Sequential-Forward-Selection (SFS) procedure [Devijver, 82], based on the performance measured from the Caterpillar User Manual, is then adopted to rank different features. Among them, the Chinese transfer lexicon feature (abbreviated as CTL in the table), which only adopts Normalized-Transfer-Lexicon-Matching-Measure and matching-type priori distribution (i.e.,  $P(type_{i,j})$ ), is first selected, then CL feature (which adopts character-length), WL feature (using word-length) and EC feature (using English cognate) follow in sequence, as reported in Table 4.1.

The selection sequence verifies our previous supposition that the transfer-lexicon is a more reliable feature and contributes most to the aligning task. Table 4.1 clearly shows that the proposed robust model achieves a 60%  $F$ -measure error reduction (from 14.4% to 5.8%) compared with the baseline model (i.e., improving the cross-style performance from 85.6% to 94.2% in  $F$ -measure). The

<sup>5</sup>Which is defined as  $\frac{2pr}{p+r}$ .

	Training Set [Caterpillar User Manual]	Testing Set I [Caterpillar User Manual]	Testing Set II [Sinorama Magazine]
Baseline Model	98.91	98.21	85.56
CTL	98.26	97.51	97.51
CTL+CL	99.32	98.19	89.61
CTL+CL+WL	99.61	98.83	94.07
CTL+CL+WL+EC	99.75	99.11	94.16

Table 4.1: Performance ( $F$ -measure %) of each model and SFS

result also indicates that the length-related features are still useful, even though they are relatively unreliable.

## 5 Error Analysis

In order to understand more about the behavior of the various features, we classify all errors which occurs in aligning Sinorama Magazine in Table 5.1; the error dominated by the prior distribution of matching type is called matching-type error, the error dominated by length feature is called length-type error, and the error caused from both length features and lexical-related features (either one is not dominant) is called length&lexicon-type error<sup>6</sup>.

From Table 5.1, it is found that the matching-type errors dominate in the baseline model. To investigate the matching-type error, the prior distributions of matching-types under training set [Caterpillar User Manual] and testing set II [Sinorama Magazine] are given in Table 5.2. The comparison clearly shows that the matching-type distribution varies significantly across different domains, and that explains why the baseline model (which only considers length-based features and matching-type distribution) fails to achieve the similar performance in the cross-style test. However, as the “1-1” matching-type always dominates in both texts, the matching-type distribution still provide useful information for aligning sentences when it is jointly considered with the lexical-related feature. For those Length-Type errors generated from the baseline model in Table 5.1, different statistical characteristics across different styles are listed in Table

<sup>6</sup>In our experiment, we do not find any error dominated by lexical-related feature.

5.3. It also clearly shows that the associated statistical characteristics of those length-based features vary significantly across different styles. Furthermore, although English-cognates are reliable cues for aligning bi-lingual sentences and occurs quite a few times in the technical manual (such as company names: IBM, HP, etc., and some special technical terms such as “RS-232”, etc), they almost never occur in a general magazine such as the one that we test. Therefore, they provide no help for aligning corpus in such domains.

Table 5.1 also shows that errors distribute differently in the proposed robust model. The length-type, instead of matching-type, now dominates errors, which implies that the mismatching effect resulting from different distributions of matching types has been diluted by the transfer-lexicon feature. Furthermore, the score of erroneous lexicon-type assignment never dominates any error found in the proposed robust model, which verifies our supposition that transfer-lexicons are more reliable cues for aligning sentences.

To further investigate those remaining errors generated from the proposed robust model, two error examples are given in Figure 1. The first case shows an example of “Length-Type Error”, in which the short sentence (E2) is erroneously merged with the long sentence (E1) and results in an erroneous alignment [E1, E2 : C1] and [E3 : C2]. (The correct alignment should be [E1 : C1] and [E2, E3 : C2].) Generally speaking, if a short source sentence is enclosed by two long source sentences in both sides, and they are jointly translated into two long target sentences, then it is error prone compared with other cases. The main reason is that this short source sentence would contain only a few words and thus its associated transfer-

Proposed Robust Model		Baseline Mode	
Error Type	Percentage (%)	Error Type	Percentage (%)
Matching-Type Error	21.9	Matching-Type Error	81.2
Length-Type Error	62.5	Length-Type Error	14.9
Length&Lexicon-Type Error	15.6	Length&Lexicon-Type Error	3.9

Table 5.1: Error Classification while aligning Sinorama Magazine

	1-0	0-1	1-1	1-2	2-1	2-2	1-3	3-1	1-4	4-1	4-2
Caterpillar	0.1%	0.25%	93.58%	2.31%	3.4%	0.05%	0.11%	0.08%	0.06%	0	0
Sinorama	0.28%	0	65.54%	1.69%	24.86%	0.28%	0	5.08%	0	1.98%	0.28%

Table 5.2: Comparison of prior distributions

	Cognate	Length Features ( $\delta_c, \delta_w$ )					$\delta_{cognate}$	$\delta_{Transfer-Lexicon}$	
	Occurrence Rate <sup>7</sup>	$c$	$s_c^2$	$w$	$s_w^2$	$r$	$\lambda$	$\mu$	$\sigma^2$
Caterpillar	36.4%	0.65	0.87	3.45	6.09	-0.02	0.06	-0.72	0.21
Sinorama	1.1%	0.59	1.79	2.76	7.80	-0.46	0.25	-0.60	0.02

Table 5.3: List of all associated parameters

lexicons are not sufficient enough to override the wrong preference given by the length-based feature (which would assign similar score to both merge-directions).

The second case shows an example of “Length&Lexicon-Type Error”, in which the source sentence (E1) is erroneously deleted and results in an erroneous alignment [E1: Delete] and [E2 : C1]. (The correct alignment should be [E1, E2 : C1].) The main reason is that the meaning of sentence (E1) is similar to that of (E2) but stated in different words, and the translator has merged the redundant information in his/her translation. Therefore, the length-feature prefers to delete the first source sentence. On the other hand, since most of those associated transfer-lexicons in the source sentence E1 cannot be found in the corresponding target sentence C1, the Transfer-Lexicon feature also prefers to delete the first source sentence E1. It seems that this kind of errors would require further knowledge from language understanding to solve them, and is beyond the scope of this paper.

<sup>7</sup>The occurrence rate is defined as “Number of sentences that contained cognates” / “Total number of sentences”

## 6 Conclusions

Although those length-based approaches are simple and can achieve good performance when they are trained and tested in the corpora of the same style, the performance drops significantly when they are tested in different styles other than that of the training corpora. (For instance, the  $F$ -measure error increases from 1.8% to 14.4% in our experiment.) The main reason is that the statistical characteristics of those features adopted by the length-based approaches (such as length-distribution, alignment-type-distribution and cognate-frequency) vary significantly from one style to another style.

Since human align sentences mainly by examining the similarity between different meanings conveyed by the given bilingual sentences pair, not by counting the number of characters in sentences, the transfer-lexicon is expected to be the more reliable cue than the sentence length. A robust statistical sentences alignment model, which integrates the associated transfer-lexicons into the original length-based model, is thus proposed in this paper. Great improvement has been observed in our experiment, which reduces the  $F$ -measure error generated from

the length-based model from 14.4% to 5.8%, when the proposed approach is tested in the cross-style case.

Last, length-features, cognate-feature and transfer-lexicon-feature are implicitly assumed to contribute equally in aligning sentences in this paper; however this assumption is not usually held because different features might have various dynamic ranges for their scores and thus contribute differently to discrimination power. To overcome this problem, various features would be weighted differently in the future.

### Acknowledgement

We would like to thank both Prof. Hsin-Hsi Chen and Prof. Kuang-Hwa Chen for their kindly providing us the aligned bi-lingual Sinorama Magazine for conducting the above experiment. The appreciation is also extended to our Translation Service Center for providing the bilingual Caterpillar User Manual for this study.

### References

- [Brown et al., 91] Peter F. Brown, Jennifer C. Lai, and Robert L. Mercer, (1991). "Aligning Sentences in Parallel Corpora", *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 169-176, 18-21 June 1991, UC Berkeley, California, USA.
- [Brown et al., 93] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra and Robert L. Mercer, (1993). "The Mathematics of Statistical Machine Translation: Parameter Estimation", *Computational Linguistics* 19: 263-311.
- [Chen, 93] Stanley F. Chen, (1993). "Aligning Sentences in Bilingual Corpora Using Lexical Information", *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pp. 9-16, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA.
- [Church, 93] Kenneth W. Church, (1993). "Char\_align: A Program for Aligning Parallel Texts at the Character Level", *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pp.1-8, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA.
- [Dagon et al., 91] Ido Dagon, Alon Itai and Ulrike Schwab, (1991). "Two Language Are More Informative Than One", *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp. 130-137, 18-21 June 1991, UC Berkeley, California, USA.
- [Devijver 82] Pierre A. Devijver and Josef Kittler, (1982). *Pattern Recognition: A Statistical Approach*, Prentice-Hall Inc., N.J., USA, 1982.
- [Gale and Church, 93] William A. Gale and Kenneth W. Church, (1991). "A Program for Aligning Sentences in Bilingual Corpora", *Computational Linguistics* 19:75-102.
- [Och and Ney, 2000] Franz Josef Och and Hermann Ney, (2000). "A Comparison of Alignment Models for Statistical Machine Translation", *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 1086-1090, 1-8 October 2000, Hong Kong.
- [Rabiner and Juang, 93] Lawrence Rabiner and B.H. Juang, (1993). *Fundamentals of Speech Recognition*, Prentice-Hall Inc., N.J., USA, 1993.
- [Su et al., 95] K. Y. Su, J. S. Chang and Una Hsu, (1995). "A Corpus-Based Statistics-Oriented Two-Way Design", *Proceedings of TMI-95*, Vol. II, pp. 334-353, Centre for Computational Linguistics, Katholieke Universiteit Leuven, Leuven, Leuven, Belgium, July 5-7, 1995.
- [Su and Chang, 99] K. Y. Su and J. S. Chang, (1999). "A Customizable, Self-Learnable Parameterized MT System: Text Generation", *Proceedings of MT SUMMIT VII*, pp. 182-188, Singapore. (Invited Talk)
- [Vogel et al., 96] Stephan Vogel, Hermann Ney and Christoph Tillmann, (1996). "HMM-Based Word Alignment in Statistical Translation", *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, pp. 836-841, 24-27 June 1996, UC Santa Cruz, California, USA.
- [Wu, 94] Dekai Wu, (1994). "Aligning a Parallel English-Chinese Corpus Statistically with Lexical Criteria", *Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics*, pp. 80-87, 27-30 June 1994, New Mexico State University, Las Cruces, New Mexico, USA.