Are Unaligned Words Important for Machine Translation ?

Yuqi Zhang **Evgeny Matusov** **Hermann Ney**

Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6 – Computer Science Department
RWTH Aachen University, D-52056 Aachen, Germany
{yzhang,matusov,ney}@cs.rwth-aachen.de

Abstract

In this paper, we deal with the problem of a large number of unaligned words in automatically learned word alignments for machine translation (MT). These unaligned words are the reason for ambiguous phrase pairs extracted by a statistical phrase-based MT system. In translation, this phrase ambiguity causes deletion and insertion errors. We present hard and optional deletion approaches to remove the unaligned words in the source language sentences. Improvements in translation quality are achieved both on large and small vocabulary tasks with the presented methods.

1 Introduction

Word alignment is a key part in the training of a statistical MT system because it provides mappings of words between each source sentence and its target language translation. Because of the difference in the structure of the involved languages, not all words in the source language have a corresponding word in the target language. So in the alignments, no matter manually created or automatically learned, some words are aligned, some are not.

Current state-of-the-art statistical machine translation is based on phrases. First the word alignments for the training corpus are generated. Then phrase alignments are inferred heuristically from the word alignments. This approach was presented by (Och et al., 1999) and implemented by e. g. (Koehn et al., 2003). Since this widely used phrase extraction method depends on word alignments, it is often assumed that the quality of word alignment is critical to the success of translation. However, some research have shown that the large gains in alignment accuracy often lead to, at best, minor gains in translation performance (Lopez and Resnik, 2006). They concluded that it could be more useful to directly investigate ways to reduce the noise in phrase extraction than improving word alignment. The work by (Ma et al., 2007) shows that a good phrase segmentation is important for translation result. Encouraged by the work, this paper explores the influence of the unaligned words on the phrase extraction and machine translation results. We show that the presence of unaligned words causes extraction of "noisy" phrases which can lead to insertion and deletion errors in the translation output. Furthermore, we propose approaches for "hard" and "soft" deletion of the unaligned words on the source language side. We then show that better way to deal with unaligned words can substantially improve translation quality, on both small and large vocabulary tasks.

In section 2, we briefly review the word alignment concept and point out that there is a large number of unaligned words in both manual and automatic alignments used for common translation tasks. In section 3, we explain how the unaligned words affect the phrase extraction and cause deletion and insertion errors. In section 4, we present two approaches to prove the negative impact of the unaligned words on translation quality. The experimental results are given in sections 5 and 6. Finally, section 7 presents a conclusion and future work.

2 Unaligned words in word alignment

In statistical translation models (Brown et al., 1990), a "hidden" alignment

Proceedings of the 13th Annual Conference of the EAMT, pages 226–233,
Barcelona, May 2009

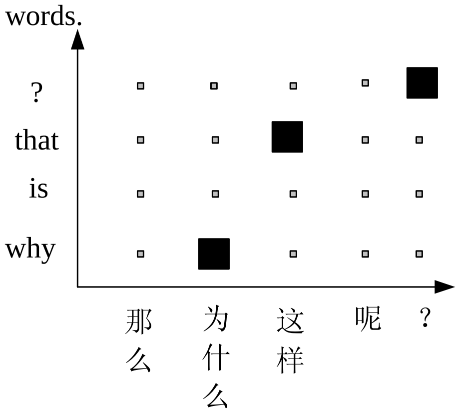
$a_1^J := a_1, \ldots, a_j, \ldots, a_J$ is introduced for aligning the source sentence f_1^J to the target sentence e_1^I. The source word at position j is aligned to the target word at position $i = a_j$. The alignment a_1^J may contain special alignment $a_j = 0$, which means that the source word at index j is not aligned to any target word. Because a word in the source sentence cannot be aligned to multiple words in the target sentence, the alignment is trained in both translation directions: source to target and target to source. For each direction, a Viterbi alignment (Brown et al., 1993) is computed: $A_1 = \{(a_j, j)|a_j \geq 0\}$ and $A_2 = \{(i, b_i)|b_i \geq 0\}$. Here, a_1^J is the alignment from the source language to the target language and b_1^I is the alignment from the target language to the source language. To obtain more symmetrized alignments, A_1 and A_2 can be combined into one alignment matrix A with the following combination methods. More details are described in (Och and Ney, 2004):

- $intersect$: $A = A_1 \cap A_2$
- $union$: $A = A_1 \cup A_2$
- $refined$: extend from the intersection. $intersect \subseteq refined \subseteq union$

In any of the alignments, there are many words which are unaligned. We have counted unaligned words in various Chinese-English alignments both a small corpus (LDC2006E93[1]) and a large corpus (GALE-All[2]). Table 1 presents what percentage of unaligned words occurs in each alignment. Since the released LDC2006E93 corpus contains manual alignments, we can see that even in "correct" alignments, more than 10% words are unaligned. $intersect$, the alignment with the best precision, has around 50% unaligned words on both sides. IN $union$, which has best recall, still around 10% of the words are unaligned. The most often used $refined$ alignment, which has the balance between precision and recall, has about 25% unaligned words. Since phrase pairs are extracted from the word alignments, these unaligned words will affect them as described below.

[1]LDC2006E93: LDC GALE Y1 Q4 Release - Word Alignment V1.0, Linguistic Data Consortium (LDC)
[2]GALE-ALL: all available training data for Chinese-English translation released by LDC. http://projects.ldc.upenn.edu/gale/data/DataMatrix.html

Figure 1: An alignment example with unaligned words.



为什么	why
为什么	why is
那么 为什么	why
那么 为什么	why is
为什么 这样	why is that
为什么 这样 呢	why is that
那么 为什么 这样	why is that
那么 为什么 这样 呢	why is that
为什么 这样 呢 ?	why is that ?
那么 为什么 这样 呢 ?	why is that ?
那么	is
呢	is
这样	is that
这样	that
这样 呢	is that
这样 呢	that
这样 呢 ?	is that ?
这样 呢 ?	that ?
呢 ?	?
?	?

3 Phrase extraction

In the state-of-the-art statistical phrase-based models, the unit of translation is any contiguous sequence of words, which is called a phrase. The phrase extraction task is to find all bilingual phrases in the training data which are consistent with the word alignment. This means that all words within the source phrase must be aligned only with the words of the target phrase; likewise, the words of the target phrase must be aligned only with the words of the source phrase (Och et al., 1999) (Zens et al., 2002). A target phrase can have multiple consistent source phrases if there are unaligned words at the boundary of the source phrase and vice versa.

Figure 1 gives an alignment example with un-

Corpus	Sentence	Alignment	Unaligned Chinese words	Unaligned English words
LDC2006E93	10,565	$manual$	14%	11%
		$intersect$	53%	40%
		$refined$	23%	23%
		$union$	7%	14%
GALE-All	8,778,755	$intersect$	48%	55%
		$refined$	24%	27%
		$union$	9%	16%

Table 1: The percentages of unaligned words in variant alignments.

aligned words on both source and target sides and the phrase table extracted from this alignment. The unaligned words will result in multiple extracted phrase pairs. All of these phrase pairs are kept because the unaligned words are necessary to complete a good sentence though they have no corresponding translations. However, the translation models are not powerful enough to select the correct phrase pair from these multiple pairs. As a result, this ambiguity often causes insertion errors which is adding redundant words to the translation and deletion errors which means that translations of some source words are missing. We have used the phrase table in figure 1 to translate the source sentence. (The translation system will be described in the section 5). Since the example sentence is short, to see how the phrase pairs are concatenated, we limit the length of the used phrase from 1 to 4. In the table 2 there is an insertion error with $slen = 1$, $tlen = 1$, which is caused by the unaligned 'is' in the phrase '呢# is'. With $slen = 2$, $tlen = 2$ and $slen = 3$, $tlen = 3$ there are deletion errors where unaligned 'is' is missing in phrase '那么 为什么why is# is'.

4 Deletion of the unaligned words in source sentences

Based on the observations in the last section, we are going to disambiguate the multiple phrase pairs caused by unaligned words. In the automatically trained alignment there are a few possible cases for the unaligned words.

correct vs. wrong: an unaligned word is correct if it really has no corresponding translations and is left unaligned by a human annotator. An unaligned word is wrong if it has been aligned in the manual alignment.

function words vs. content words: Comparing the alignment of function words and content words, we could find that the correct unaligned words are roughly function words, while the wrong unaligned words are usually content words. The function words have little lexical meaning, but instead serve to express grammatical relationships with other words within a sentence On the contrary, the content words usually carry meaning, which are "natural units" of translation between languages.

If we just focus on the disambiguation of multiple phrases and not consider applying grammatical information in function words to the translation system, like the work done by (Setiawan et al., 2007), the simplest way of reducing the multiple phrases is to delete the 'correct' unaligned words: the function words. The function words at the target side should not be touched, since they are necessary to complete a good sentence. However, the function words at the source side could be removed, when they have no corresponding translations.

4.1 Deletion Candidates

Not all unaligned words should be removed. Besides the content words, a source function word could also have correct mappings to the target words in some sentences. We have used two constraints to filter out the words which can be deleted..

We use relative frequencies to estimate the probability of a word being aligned.

$$p(w_{align}) = \frac{N_{w_{align}}}{N(w)} \tag{1}$$

The number of times a word w is aligned in the training data is denoted by $N_{w_{align}}$, and $N(w)$ is the total number of occurrences of the word w. The

slen=1 tlen=1	why # 为什么 ## is # 那么 ## that # 这样 ## is # 呢 ## ? # ？	why is that is ?
slen=2 tlen=2	why # 那么 为什么 ## that # 这样 ## ? # 呢 ？	why that ?
slen=3 tlen=3	why # 那么 为什么 ## that ? # 这样 呢 ？	why that ?
slen=4 tlen=4	why is that # 那么 为什么 这样 ## ? # 呢 ？	why is that ?

Table 2: The translations of the example with phrase length limitation. The symbol ## denotes concatenation of phrase pairs.

first constraint is that the probability of a word being aligned is below a threshold τ.

$$Con\_p(w) = \begin{cases} 1 & \text{if } p(w_{align}) \leq \tau \\ 0 & \text{if } p(w_{align}) > \tau \end{cases} \quad (2)$$

This constraint can be used with different thresholds. The smaller the threshold is, the more strict constraint is applied and fewer words are to be considered. When $p(w_{align})$ is 0.5, it means that the word has the same probability to be aligned and not to be aligned. In order to filter out the deletion candidates, the best threshold as determined in our experiments should be less than 0.5.

The second constraint is to use the POS tags to mark the function words. In general, the content words include nouns, verbs, adjectives, and most adverbs. We denote the POS tag set for content words as $S = \{noun, verb, adj, adv\}$. The constraint for the function word is:

$$Con\_fun(w) = \begin{cases} 1 & \text{if } POS(w) \notin S \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

In the experiments, we will test both $Con\_p(w)$ and $Con\_p(w)$+$Con\_fun(w)$. We will show that it is more important for a deletion candidate to be constrained by $Con\_p(w)$, since content and function words in linguistics are not always distinguished clearly.

4.2 Hard deletion

The simplest way of deletion is directly removing the found words from the source sentences and the alignments. The change of the alignment will affect not only the extracted phrase pairs around the deleted word, but also the probability estimation of all phrases. In this way, the source sentences become relatively shorter. The size of the phrase table will be smaller because of the reduction in the multiple translation pairs. However, the drawback of the method is obvious. Most words are aligned or not in different contexts. When we set τ greater than 0 and delete the filtered words, there must be some words which should actually be translated, which means that they were deleted wrongly.

Hard deletion is an easy method to investigate the influence of unaligned words on translation results. Although the method will cause overdeletion, it can reflect which multiple translation pairs containing an unaligned word provide more useful information or more harmful information for ultimate translation quality.

4.3 Optional deletion

A better and more complicated method is to apply optional deletion. We do not make a firm decision to delete any words. Instead, we preserve ambiguity and defer the decision until later stages.

We use a confusion network (CN) to represent the ambiguous input. Some works are reported to use CNs in machine translation (Bertoldi et al., 2007; Koehn et al., 2007). A CN is a directed acyclic graph in which each path goes through all the nodes from the start node to the end node. Its edges are labeled with words. An example of a CN for optional deletion is shown in table 3.

把.38	机票1.0	忘1.0	在1.0	家里1.0	了.21
ε.62					ε.79

Table 3: A CN example of optional deletion.

The special empty-word ε represents a word deletion. Also, the word aligned probability is attached to each edge. The probability is calculated by equation (1). When the word is a content word, its aligned probability is 1.0. The score with $epsilon$ means the probability of the word in the same column not to be aligned, which is equal to $1 - p(w_{align})$.

Input source sentences are represented by CN. Like what is done in the hard deletion, the alignments are modified by deleting all deletion candidates and the corresponding points in the alignment matrix. However, to match the possible non-deletion of the unaligned words, the original alignment is also needed. We combine the two alignments by merging the phrase counts and recompute the phrase probabilities.

5 Experimental Setup

5.1 Data

We carried out MT experiments for translation from Chinese to English on two data sets: BTEC08 and GALE08.

The BTEC08 data was provided within the IWSLT 2008 evaluation campaign (Paul, 2008), extracted from the Basic Traveling Expression Corpus(BTEC) (Takezawa et al., 2002). The data is a multilingual speech corpus which contains sentences which are usually found in books for tourists. The sentences are short, with less than 10 words on average. The parallel training data is relatively small. We added the official IWSLT08 training data, the IWSLT06 dev data and IWSLT06 evaluation data and their references to the training data. The development and test sets in the experiments below are from the IWSLT04 and IWSLT05 evaluation data. We found that the two data sets are not similar, so we took the first half of each and combine them as dev data. The remaining two halves are combined as test data.

The large vocabulary GALE data were all provided by LDC. The test data has four genres: broadcast news (BN), broadcast conversations (BC), newswire (NW) and webtext (WT). The first two genres are for speech translation and the last two are for text translation. Here, we only carried out experiments on NW. The sentences of the GALE task are longer (around 30 words per sentence) and more difficult to translate.

The corpus statistics for both tasks are shown in Table 4:

BTEC		Chinese	English
Train:	Sentences	23940	
	Running words	181486	232746
Dev:	Sentences	503	
	Running words	3085	3887
Test:	Sentences	503	
	Running words	3109	3991
GALE		Chinese	English
Train:	Sentences	8778755	
	Running words	232799466	249514713
Dev08:	Sentences	485	
	Running words	14750	16570
Test08:	Sentences	480	
	Running words	14800	16683

Table 4: Corpus Statistics of the BTEC and GALE translation tasks. For BTEC dev and test sets, the number of English tokens is the average over 16 human reference translations.

5.2 Baseline System

Our baseline system is a standard phrase-based SMT system. Word alignments are obtained by using GIZA++ (Och and Ney, 2003) with IBM model 4 [3]. We symmetrized bidirectional alignments using the refined heuristic (Och and Ney, 2004). The phrase-based translation model is a log-linear model that include phrase translation probabilities and word-based translation probabilities in both translation directions, phrase count models, word and phrase penalty, target language model (LM) and a distortion model. Language models were built using the SRI language modeling toolkit (Stolcke, 2002). On the small vocabulary BTEC task we used a 6-gram. On the large vocabulary GALE task we included 5-gram language model probabilities. The model scaling factors are optimized on the development set with the goal of improving the BLEU score. We used a non-monotonic phrase-based search algorithm that can take confusion networks as input in the style of (Matusov et al., 2008).

[3]Specifically, on GALE data we performed 5 iterations of Model 1, 5 iterations of HMM, 2 iterations of Model 4. On BTEC data we performed 4 iterations of Model 1, 5 iterations of HMM, 8 iterations of Model 4.

6 Experimental Results

6.1 The deletion candidates

First, we tested different thresholds τ in the range from 0.2 to 0.5. The set with small τ is a subset of the one with larget τ. We filtered out words which were most frequently not aligned in training. We performed the experiments on both BTEC and GALE tasks. The findings are reported in table 5. For each threshold the table gives the number of unique words removed (num.) and some examples.

By applying the two constraints Con_p+Con_fun, the number of deletion candidates is reduced greatly. That means among the unaligned words in alignments there are many content words. The content words, especially nouns, usually are expected to be translated. It is not good if there are many content words unaligned.

Comparing between BTEC and GALE there are fewer deletion candidates in GALE data, both con-

	BTEC				GALE			
	Con_p		Con_p+Con_fun		Con_p		Con_p+Con_fun	
τ	num.	example	num.	example	num.	example	num.	example
0.2	1	的	1	的	1	恭	0	-
0.3	4	的 了 哭 却	3	的 了 却	7	的 慧 毛病 . . .	1	的
0.4	21	叶 以 把 战争 . . .	10	以 把 . . .	17	的 罗斯 . . .	1	的
0.5	152	呀 对 着 当时 . . .	20	呀 对 着 . . .	62	的 中 之 兆. . .	3	的 中 之

Table 5: Some statistics and examples of the words removed based on the constraints defined in equations 2 and 3.

tent and function words. It implies that large data leads to obtain better alignments which assign more mappings between source and target languages.

6.2 Hard deletion

Since the hard deletion is easy to carry out, we performed the experiments on both BTEC and GALE tasks here, too. As the number of deletion candidates on GALE is small, we tested the smallest deletion candidate set "的" and the biggest set which is under the constraint Con_p with $\tau = 0.5$. Translation results are shown in table 6. The second row "rm-1" is the hard deletion of 的 and the third row "rm-62" is for the deletion of the 62 words as shown in table 5.

It is interesting to see that the deletions of both the small set and large set of words improve the baseline on every metric. 的 is the most common function word in Chinese to connect adjectives and nouns and it is also the word with lowest aligned probability in the table 5. The BLEU and TER scores both improve 0.5% absolute on dev and test data just by removing this single word. However, when we remove the 62 words including 的, the result does not improve further. This means that the deletion candidate set contains some content words, the deletion of which has a negative influence on translation quality.

The BTEC data provides us with a larger deletion candidate set. Additionally, the small size of the training data for the BTEC task makes it possible to run some finer-grained experiments. We focus on how the removable function words affect the translation quality. The experiments are carried on the word set with different thresholds τ and under the constraint Con_p+Con_fun. The translation results with hard deletion on BTEC are shown in table 7.

The improvement in the BTEC data is not as much as on the GALE data. Only when τ is set to 0.4, we obtained slightly better scores. The reason is that extracted phrases are very long comparing to the sentence length. The maximum phrase length was set to 15 words, both for BTEC and GALE task. However, the average sentence length of the BTEC test set is around 7 words, vs. 30 words on the GALE task. When phrase pairs are longer, there are fewer cases that unaligned words are at their boundaries. The translation examples in table 2 also reflect this phenomenon. That source sentence has 5 words. When the phrase length limitation is 4, unaligned 'is' is an inner word in the phrase pair *why is that* # 那么 为什么 这样.

6.3 Optional deletion

In addition to the hard deletion experiments on BTEC, we carried out the optional deletion experiments in the same settings. The results are also shown in table 7. The optional deletion method achieved good performance. The BLEU score improves consistently with all settings, at most 1.5% on the dev set and 0.7% on the test set with $\tau = 0.4$.

Furthermore, we are also interested in the influence of individual deletion candidate on the translation results. It would be more useful if we know what words are important for the deletion instead of just determining the optimal threshold. Since $\tau = 0.4$ has achieved the best result both in hard deletion and optional deletion, we explore the 10 removable function words in the set one by one. The 10 words are listed in table 8. At first, we sorted the 10 words according to the probability of being aligned. From the low to high probability, we add one word a time to the deletion candidate set. The results are shown in table 8. The word 的, which has the lowest probability of being aligned, is the most important word in the set.

	dev08				test08			
%	BLEU	Interval	TER	Interval	BLEU	Interval	TER	Interval
baseline	31.5	[30.4, 32.7]	60.7	[59.9, 61.5]	30.9	[29.8, 32.0]	60.3	[59.5, 61.1]
rm-1:的	31.9	[30.6, 33.1]	60.2	[59.1, 61.2]	31.4	[30.3, 32.7]	59.7	[58.9, 60.7]
rm-62	32.3	[31.0, 33.6]	60.1	[59.0, 61.0]	31.2	[30.0, 32.3]	59.9	[59.0, 60.8]

Table 6: Translation results using the hard deletion method on the GALE task.

	dev				test			
%	BLEU	Interval	TER	Interval	BLEU	Interval	TER	Interval
baseline	49.6	[47.0, 52.6]	41.3	[39.1, 43.5]	49.5	[46.8, 52.1]	41.3	[39.3, 43.3]
rm-funW	Hard deletion							
$\tau = 0.2$	49.1	[46.3, 52.1]	41.9	[39.6, 43.9]	49.7	[47.0, 52.3]	41.5	[39.5, 43.6]
$\tau = 0.3$	50.0	[47.1, 52.9]	41.0	[38.8, 43.5]	49.3	[46.4, 51.9]	41.2	[39.3, 43.6]
$\tau = 0.4$	50.0	[46.9, 52.9]	41.3	[39.4, 43.8]	49.7	[47.1, 52.6]	41.1	[39.0, 43.3]
rm-funW	Optional deletion							
$\tau = 0.2$	51.1	[48.6, 54.2]	40.5	[38.2, 42.7]	49.6	[46.7, 52.6]	41.5	[39.3, 43.7]
$\tau = 0.3$	51.2	[48.9, 53.9]	40.4	[38.5, 42.1]	49.9	[47.1, 52.6]	41.5	[39.5, 43.8]
$\tau = 0.4$	51.1	[48.5, 53.5]	40.6	[38.7, 42.9]	50.2	[47.7, 53.0]	41.4	[39.3, 43.5]

Table 7: Translation results using hard and optional deletion methods on the BTEC task.

We also calculate the 95% confidence intervals for both hard deletion and optional deletion. Unfortunately, the new systems are not statistical significant though the BLEU scores are better.

7 Conclusion and future work

In this paper, we have devoted attention to the problem of a large number of unaligned words in the word alignments generally used for MT model training. These unaligned words result in ambiguous phrase pairs being extracted by a state-of-the-art phrase-based statistical MT system. In translation, this phrase ambiguity causes deletion and insertion errors. We classified the unaligned words into function words and content words and showed that unaligned function words have an important influence on phrase extraction.

Furthermore, we have proposed two methods to improve phrase extraction based on handling of unaligned words. Since it is important to keep the unaligned words on the target side to obtain complete and fluent translations, we have applied hard deletion and optional deletion of the unaligned words on the source side before phrase extraction. Though the methods are simple, they still achieved notable improvements in automatic MT evaluation measures on both small and large vocabulary tasks. We have shown that differentiating between useful and “removable” unaligned words is important for the quality of the extracted phrases and, consequently, for the quality of the phrase-based MT.

This paper pointed out the importance of unaligned words, but only considered the source language words. In the future, more work should be done regarding the unaligned words in the target language. The translations are more directly affected by the quality of target phrases. Since deleting of unaligned words at the target side is clearly not the right solution, some disambiguation models are to be investigated.

8 Acknowledgments

This material is partly based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR0011-06-C-0023 and partly realized as part of the Quaero Programme, funded by OSEO, French State agency for innovation.

References

P.F.Brown, J.Cocke, S.A.Della Pietra, V.J.Della Pietra, F. Jelinek, J.D.Lafferty, R.L.Mercer, and P.S. Roossin. 1990. A statistical approach to machine translation., In *Computational Linguistics,16(2)*, pages 79-85, Jun.

P.F.Brown, S.A.Della Pietra, V.J.Della Pietra, and R.L.Mercer. 1993. The mathematics of statistical

		Hard deletion		optional deletion	
		dev	test	dev	test
	$p(w_{align})$	BLEU[%]	BLEU[%]	BLEU[%]	BLEU[%]
baseline	-	49.6	49.5	49.6	49.5
的	0.007	49.1	49.7	51.1	49.6
+ 了	0.21	50.0	49.3	51.2	49.9
+ 却	0.27	50.0	49.3	51.2	49.9
+ 以	0.35	50.0	49.4	51.2	49.9
+ 把	0.38	50.0	49.7	51.1	50.2
+ 对于	0.4	50.1	49.7	51.1	50.1
+ 既	0.4	50.1	49.6	51.1	50.1
+ 着	0.4	50.1	49.7	51.1	50.1
+ 式	0.4	50.1	49.6	51.1	50.1
+ 对	0.4	50.0	49.7	51.1	50.2

Table 8: The influence of deleting individual words on the translation quality (BTEC task).

machine translation: Parameter estimation., In *Computational Linguistics,19(2)*, pages 263-311

Nicola Bertoldi, Richard Zens and Marcello Federico. 2007. Speech Translation by Confusion Network Decoding, In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1297-1300, Apr.

Philipp Koeln, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation, In *Proceedings of the 2003 Human Language Technology onference of the North American Chapter of the Association for Computational Linguistics(HLT-NAACL)*, pages 127-133, May.

Philipp Koehn, Nicola Bertoldi, Ondrej Bojar, Chris Callison-Burch, Alexandra Constantin, Brooke Cowan, Chris Dyer, Marcello Federico, Evan Herbst, Hieu Hoang, Christine Moran, Wade Shen, and Richard Zens. 2007. Open Source Toolkit for Statistical Machine Translation: Factored Translation Models and Confusion Network Decoding, *CLSP Summer Workshop Final Report WS-2006,*

Adam Lopez and Philip Reesnik. 2006. Word-Based Alignment, Phrase-Based Translation: What's the Link?, In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas,* pages 90-99, Aug.

Evgeny Matusov and Björn Hoffmeister and Hermann Ney. 2008. ASR Word Lattice Translation with Exhaustive Reordering is Possible, In *Proceedings of the Interspeech 2008,* pages 2342-2345, Sep.

Yanjun Ma and Nicolas Stroppa and Andy Way 2007. Bootstrapping Word Alignment via Word Packing In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics,* pages 304–311, Jun. Prague, Czech Republic.

Franz Josef Och, Christoph Tillman, and Hermann Ney. 1999. Improved alignment models for statistical machine translation, In *Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methos in Natural Language Processing and very Large Corpora(EMNLP-VLC),* pages 20-28, Jun.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models, In *Computational Linguistics,29(1):19-51*

Franz Josef Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation, In *Computational Linguistics*

Michael Paul. 2008. Overview of the IWSLT 2008 evaluation campaign, In *Proceedings of the International Workshop on Spoken Language Translation,* pages 1–17, Oct.

Andrea Stolcke. 2002. SRILM - An extensible language modeling toolkit, In *Proceedings of the International Conference on Spoken Language Processing,* pages 901-904.

Hendra Setiawan and Min-Yen Kan and Haizhou Li. 2007. Ordering Phrases with Function Words, In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* pages 712–719, Jun.

Toshiyuki Takezawa and Eiichiro Sumita and Fumiaki Sugaya and Hirofumi Yamamoto and Seiichi Yamamoto. 2002. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world, In *Proceedings of Third International Conference on Language Resources and Evaluation 2002 (LREC)*, pages 147-152

Richard Zens, Franz Josef Och and Hermann Ney. 2002. Phrase-based statistical machine translation, In *25th German Conference on Artificial Inteligence (LNAI),* pages 18-32, Sep.