# Improving Sentence Segmentation Model for Statistical Machine Translation

**Zhengshan Xue　Dakun Zhang　Lina Wang　Jie Hao**
Toshiba (China) R&D Center, Beijing, China
{xuezhengshan,zhangdakun,wanglina,haojie}@toshiba.com.cn

## Abstract

The proportion of long sentences increases with the size of training corpora in statistical machine translation. How to effectively use the information in long sentences to improve the translation quality is a main challenge. This paper proposes a new method for long sentence segmentation in the training process based on (Xu et al., 2005)'s sentence segmentation model. This method can automatically get boundary words and their probabilities without manual intervention, which results more meaningful segmentation in semantics. Also, the length of segmented sub sentences are balanced through both source and target languages. Experiments on the NIST test sets show a maximum improvement of 0.5 BLEU scores.

## 1 Introduction

In recent years, Statistical Machine Translation (SMT) (Yamada and Knight, 2001; Koehn et al., 2003; Liu, 2003; Liu et al., 2006; Koehn et al., 2007; Chiang, 2007; He et al., 2008; Li et al., 2009; Xiao et al., 2012) gradually becomes a research focus in the field of Natural Language Processing (NLP). SMT relies on bilingual parallel corpora to train the parameters needed in models. Inevitably, there are lots of quite long sentences in training data. Long sentences would cost more in the training stage and the knowledge learned (such as word alignment or phrase) would be less accurate. In the actual decoding, longer sentences need more computational cost but produce worse translation due to very long distance reordering problem. Most translation systems (Koehn et al., 2007) remove long sentences with a certain length in training stage. Removing these long sentences during training will not reduce systems performance (Xu et al., 2005). That is because longer sentences have relative worse word alignment, which will influence the after processes in translation system (Meng et al., 2009). In order to effectively utilize the information in long sentences, the method of splitting long sentences into several shorter sub-sentences (i.e. sentence segmentation) is usually used.

Many researches have been made on sentence segmentation in training process. (Kim and Ehara, 1994) proposed a rule-based segmentation method, which achieved good result. But it is difficult to maintain, because writing rules manually is laborious and time consuming. (Nevado et al., 2003) used word information to search for segmentation positions by dynamic programming algorithm. However, this method also needs to collect anchor words manually and permits non-monotone alignment only. (Xu et al., 2005) proposed a sentence segmentation model based on IBM Model 1, which allowed monotone and non-monotone alignment. Integrating with length balance factor and inverse translation model, their model achieves better results in two Chinese-English translation tasks. However, whether the segmented units are meaningful were not considered in their method. Based on (Xu et al., 2005)'s method, (Meng et al., 2009) added semantic guidance and Poisson distribution ratio and improved (Xu et al., 2005)'s method effectively. Their method also needs to collect some boundary words manually. And they did not consider the probability of a same boundary word being either the beginning or ending of a sub sentence.

Other researches focus on segmenting sentences in decoding process. (Doi and Sumita, 2003) proposed to split speech output with N-gram and three evaluation standards. (Furuse et al., 1998) proposed a speech output segmentation based on semantic distance. (Sudoh et al., 2010) segmented the training set and test set into multiple sub sentences with parser, and then rewrite the sentences by introducing nonterminals. They also proposed sub sentences alignment model based on graphs. Their methods improve the reordering of long sentences and the quality of translation systems. But it relies on the quality of the parser, and when there is no subordinate relationship in these sub sentences, the effect is weakened.

In summary, most methods used in the training process need to write rules or collect anchor words manually. (Xu et al., 2005)'s method does not need manual intervention, but the segmentation positions are arbitrary which may produce some meaningless segments. (Meng et al., 2009)'s method effectively improves (Xu et al., 2005)'s sentence segmentation model by defining the boundary words. However, all words on segmentation positions are requested to be contained in the boundary word set.

To solve the above problems, this paper proposes a new segmentation method implemented automatically in training process for long sentences. This method does not rely on manually collected rules, anchor words or boundary words, which overcome the disadvantages of (Meng et al., 2009)'s method and solve (Xu et al., 2005)'s problem of arbitrary segmentation. Besides, this method is language independent and can be used for any language pairs. It works in the training process and therefore is also independent with the core engine (supports both phrase-based or syntax-based method). Four steps are included in this method: (1) Use GIZA++ to obtain word alignment and lexical probability. (2) Get hierarchical tree from word alignment by using (Zhang et al., 2008)'s SRA (Shift-Reduce-Algorithm) method. (3) Based on the output of step (2), get lexical boundary probability by maximum likelihood method. (4) Integrate the boundary words probability into (Xu et al., 2005)'s method, and calculate the possible segmentation positions. The experiment on NIST evaluation data shows an effective improvement on translation quality of SMT system.

The rest of the paper is structured as follows. Section 2 introduces (Xu et al., 2005)'s and (Meng et al., 2009)'s segmentation methods which are the basis of this paper. In section 3, we describe the long sentence segmentation model with SRA algorithm. Section 4 illustrates the experimental results and shows some examples. In the last section, we give a brief conclusion.

## 2 Segmentation Model

A bilingual sentence pair $(f,t)$, wherein $f = f_1 f_2 \dots f_{m-1} f_m$ is the source language sentence including m words, $t = t_1 t_2 \dots t_{n-1} t_n$ is the target language sentence including n words. We define a parallel segment$(f', t')$,wherein $f' = f_{j_1} f_{j_1+1} \dots f_{j_2-1} f_{j_2}$ , $t' = t_{i_1} t_{i_1+1} \dots t_{i_2-1} t_{i_2}$ , then $0 < j_1 \le j_2 \le m$ , $0 < i_1 \le i_2 \le n$. Initially, $f' = f$ and $t' = t$ and usually we segment each candidate into two pieces at a time.

### 2.1 (Xu et al., 2005)'s Segmentation Model

(Xu et al., 2005) searches for the best segmentation positions according to the IBM Word Alignment Model 1, thus split a sentence pair into two mutually independent sub sentence pairs. This requires the calculation of probability of each sub sentence pair. Assume $(j_1, i_1)$ and $(j_2, i_2)$ are the beginning and ending position of the segmented sub sentence pairs. Formula (1) calculates the probability of these sub sentence pairs, wherein $p(f_j|t_i)$ represents the lexical translation probability generated by IBM Model 1.

$$p(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = \prod_{j=j_1}^{j_2} [\frac{1}{i_2 - i_1 + 1} \sum_{i=i_1}^{i_2} p(f_j|t_i)] \quad (1)$$

Based on Formula (1), (Xu et al., 2005) incorporated two improved factors:

(1) Length balance factor.

$$p_\gamma(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = p(f_{j_1}^{j_2}|t_{i_1}^{i_2})^\gamma \quad (2)$$

$$\gamma = \beta * \frac{1}{j_2 - j_1 + 1} + (1 - \beta) \quad (3)$$

$\beta$ is a balance weight.

(2) Inverse translation model.

$$p_n(f_{j_1}^{j_2}, t_{i_1}^{i_2}) \approx p_\gamma(f_{j_1}^{j_2}|t_{i_1}^{i_2}) * p_\gamma(t_{i_1}^{i_2}|f_{j_1}^{j_2}) \quad (4)$$

Formula (1) is then revised into (4) the product of probability of direct and inverse translation.

Further, they consider two alignments between segmented sub sentence. Assume $j, i$ are arbitrary segmentation position, wherein $j \in [j_1, j_2 - 1]$, $i \in [i_1, i_2 - 1]$, two alignments are defined as:

(1) Monotone alignment $p_{j,i,1}$:

$$p_{j,i,1}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = p_n(f_{j_1}^{j}, t_{i_1}^{i}) * p_n(f_{j+1}^{j_2}, t_{i+1}^{i_2}) \quad (5)$$

(2) Non-Monotone alignment $p_{j,i,0}$:

$$p_{j,i,0}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = p_n(f_{j_1}^{j}, t_{i+1}^{i_2}) * p_n(f_{j+1}^{j_2}, t_{i_1}^{i}) \quad (6)$$

Object function:

$$(j', i', \delta') = argmax\{p_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2})\}$$

$$\delta \in \{0, 1\} \quad (7)$$

Thus, we can obtain segmentation position and alignment through Formula (7).

### 2.2 (Meng et al., 2009)'s Segmentation Model

The segmentation position and alignment in (Xu et al., 2005)'s method rely on the probability of Formula (7). However, (Xu et al., 2005) didn't consider whether the segments obtained had semantic meanings or not. (Meng et al., 2009) found that, many long sentences are complex sentences which have clauses guided by conjunction words, such as "when", "which" in English. Besides conjunctions, punctuations are also clues for segmentation. These words generally indicate the beginning or ending of an integral segment, and their positions can be marked as segmentation candidates. The generated sub sentences split from these positions would be more practical in semantics. So (Meng et al., 2009) collected four word sets ($WL_{f,s}$ and $WL_{t,s}$ list the beginning words of sub sentences or segments for source language and target language, $WL_{f,e}$ and $WL_{t,e}$ list the ending words of sub sentences or segments for source language and target language).

If the words at segmented position $(j, i)$ meet the following formula:

$$f_{j+1} \in WL_{f,s} \quad (8)$$

$$f_j \in WL_{f,e} \quad (9)$$

$$t_{i+1} \in WL_{t,s} \quad (10)$$

$$t_i \in WL_{t,e} \quad (11)$$

then

$$p_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = p_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) * (2 - p_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2}))$$

$$\delta \in \{0, 1\} \quad (12)$$

Based on (Xu et al., 2005)'s method, (Meng et al., 2009)'s method is more practical in semantics. The experiment showed (Meng et al., 2009)'s method achieved better results.

## 3 Improved Sentence Segmentation Model

### 3.1 Automatic Extracting of Boundary Words

(Xiong et al., 2010) obtained the hierarchical structure tree of bilingual sentence pairs by using bilingual word aligned corpus and SRA (Zhang et al., 2008) algorithm. Each node in the structure tree is defined as a translation zone, which includes multiple words. By defining the leading word and tailing word in the translation zone, (Xiong et al., 2010) got the training corpus with word labels. Inspired by (Xiong et al., 2010)'s method, we can get the boundary word set of segmentation positions of long sentences automatically using the same method. Besides the boundary words of source language, we also obtain the boundary words of target language. An example will be illustrated referring (Xiong et al., 2010)'s sentence.

Figure 1a shows an example of many-to-many alignment, Figure 1b is the tree representation of the word alignment after hierarchical analysis using SRA. Each node in Figure 1b is a bilingual phrase pair. Each node contains boundary words marking the beginning and ending of source language and target language of the phrase pair (the node with length 1 will not be considered due to the same beginning and ending boundary word). After traversing nodes with length longer than 1, we got the following boundary word sets:

$WL_{f,s}$ = {过去,五,因故}
$WL_{f,e}$ = {次,飞行,都,失败}
$WL_{t,s}$ = {The,due,failed}
$WL_{t,e}$ = {last,five,flights,accidents,all}

### 3.2 Calculation of Boundary Word Probability

Four boundary word sets can be obtained using SRA algorithm. Since a certain word may be either the beginning or ending of a boundary in different contexts. There may be overlaps between $WL_{f,s}$ and $WL_{f,e}$, $WL_{t,s}$ and $WL_{t,e}$. We can define the probability $p_b(w, \delta, \theta)$ of each boundary word $w$, which demonstrates the probability of being the beginning and ending of a boundary.

$$p_b(w, \delta, \theta) = \frac{count(w \in WL_{\delta,\theta})}{\sum_{\theta \in (s,e)} count(w \in WL_{\delta,\theta})}$$

$$\delta \in (f, t) \quad \theta \in (s, e) \quad (13)$$

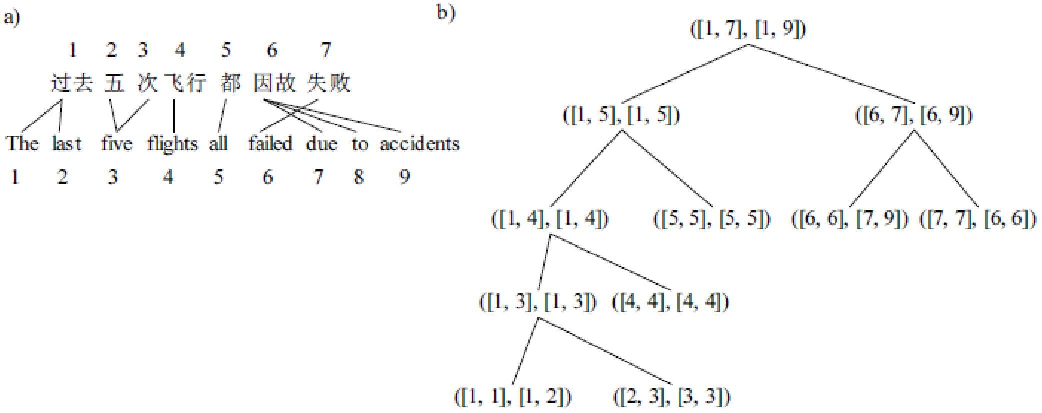

Figure 1: An example of many-to many word alignment and its tree representation produced by (Zhang et al., 2008)'s shift-reduce algorithm

wherein $f$ refers to the source language, $t$ refers to the target language, $s$ refers to the beginning of a boundary, $e$ refers to the ending of a boundary.

### 3.3 Our Segmentation Method

Based on (Xu et al., 2005)'s method, we add one factor $v(f,t,j,i)$ to calculate whether the segmentation position is meaningful. The factors $r_1$ and $r_2$ are set to balance the length ratio of the segmented sub sentence pairs from different views of source and target language.

We use same notations as in section 2. Assume $(j,i)$ are arbitrary segmentation position, and $v(f,t,j,i)$ is defined as:

$$v(f,t,j,i) = (1+p_b(f_j,f,e))^2 * (1+p_b(f_{j+1},f,s))^2$$

$$*(1+p_b(t_i,t,e))^2 * (1+p_b(t_{i+1},t,s))^2 \quad (14)$$

If $f_j$, $f_{j+1}$, $t_i$ and $t_{i+1}$ are in the boundary word sets, which means these boundary words will again act as they are, thus a bonus score will be given.

We further define variables $r_1$, $r_2$ to balance the length ratio of the segmented sub sentences pairs. Similarly, we consider two alignments:

(1) Monotone alignment:

$$r_1 = \frac{min(j-j_1+1, i-i_1+1)}{max(j-j_1+1, i-i_1+1)} \quad (15)$$

$$r_2 = \frac{min(j_2-j, i_2-i)}{max(j_2-j, i_2-i)} \quad (16)$$

(2) Non-Monotone alignment:

$$r_1 = \frac{min(j-j_1+1, i_2-i)}{max(j-j_1+1, i_2-i)} \quad (17)$$

$$r_2 = \frac{min(j_2-j, i-i_1+1)}{max(j_2-j, i-i_1+1)} \quad (18)$$

Then we revise the previous formula as:

$$p'_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) = p_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2}) * v(f,t,j,i)$$

$$*r_1 * r_2 \quad (19)$$

and object function:

$$(j',i',\delta') = argmax\{p'_{j,i,\delta}(f_{j_1}^{j_2}|t_{i_1}^{i_2})\}$$

$$\delta \in \{0,1\} \quad (20)$$

According to Formula (19), meaningful segmentation position will be awarded. If the segmented sub sentence length is not balanced, the segmentation probability from this position will be punished. For example, suppose the sentence length ratio between source and target language is 27:20 before splitting and the segmentation position is (24,2). For monotone alignment, $r_1 = 1/12$ and $r_2 = 1/6$. For non-monotone alignment, $r_1 = 3/4$ and $r_2 = 2/3$. That is, non-monotone alignment in this circumstance will be preferred. In practice, we further restrict this ratio to no more than 5. If there are segmented candidates whose length ratio is bigger than 5, we discard this candidates in our experiments.

Similar to (Xu et al., 2005)'s method, we use Formula (20) to recursively calculate the segmentation position. This procedure will continue until there is no possible segmentation position.

### 3.4 Dynamic Parameter

Four parameters are set for controlling the sentence segmentation process.

(1) ClobalMaxLen and GlobalMinLen:

ClobalMaxLen and GlobalMinLen are set in the scope of all sentences. GlobalMaxLen is used to check whether the sentence need to be split (if the length of bilingual sentence pairs is no more than GlobalMaxLen, we do not split the sentences, and otherwise, we should split them.). GlobalMinLen is used to check the length of minimum sub sentences (segments) after segmentation.

(2) LocalMaxLen and LocalMinLen:

LocalMaxLen and LocalMinLen are set for each sentence which needs to be segmented and they are dynamically changed according to the following rules.

a) Split the sentence with punctuations and get pre-segmented candidates. minLen($\tilde{f}$), maxLen($\tilde{f}$), minLen($\tilde{t}$) and maxLen($\tilde{t}$) represent the minimum or maximum length of each pre-segmented candidates $\tilde{f}$ and $\tilde{t}$ of source and target language.

b) LocalMaxLen=max(maxLen($\tilde{f}$),maxLen($\tilde{t}$));
LocalMinLen=max(minLen($\tilde{f}$),minLen($\tilde{t}$));

c) LocalMaxLen=GlobalMaxLen
if LocalMaxLen $>$ GlobalMaxLen;
LocalMinLen=GlobalMinLen
if LocalMinLen $<$ GlobalMinLen;

We restrict the length of segmented sub sentences between LocalMinLen and LocalMaxLen. The reason lies that, during the experiments, we found the length of split sub sentences is close to the value of GlobalMinLen, which results in many too short segmentations (GlobalMinLen is generally set as 1 or 2). Therefore, we can get more balanced segments under the help of LocalMinLen and LocalMaxLen.

## 4 Experiment

We made experiments translating from Chinese to English to evaluate the effectiveness of this segmentation method. We use open-source toolkit Moses[1](Koehn et al., 2007) as default decoder (phrase-based model). The length of phrases is set as 7. We used LDC2005T10[2] Chinese-English corpus to train our translation model. The target language corpus of the training data was used to train the 5-gram language model. The test set of NIST2002 was used as dev set. The test sets of NIST2002-NIST2006 and NIST 2008 were used as test corpus, which had 4 translations for reference. The parameters of the sentence segmentation were set as : GlobalMaxLen=20, GlobalMinLen=1, $\beta = 0.9$. The basic information of the corpus are shown in Table 1.

| Corpus | # of Sentences |
|---|---|
| LDC2005T10(training) | 280,766 |
| NIST2002(dev/test) | 878 |
| NIST2003(test) | 918 |
| NIST2004(test) | 1,597 |
| NIST2005(test) | 1,082 |
| NIST2006(test) | 1,664 |
| NIST2008(test) | 1,357 |

Table 1: Basic information of corpus

We design three experiments. (1) Experiment with baseline system. We train the model with original data as baseline system (without sentence segmentation). (2) Experiment with (Xu et al., 2005)'s system. We segmented the training corpus with (Xu et al., 2005)'s method and train the system. (3) Experiment with our new segmentation method. The training corpus were segmented using our new segmentation method. The lexical probability used in experiment (2) and (3) were from experiment (1). The boundary probability used in experiment (3) was obtained automatically from baseline training corpus by SAR algorithm. There is no comparison between (Meng et al., 2009)'s method and ours because the manually collected boundary words adopted in their experiment could not be confirmed. Table 2 shows the experimental results evaluated by BLEU score.

| Test set | BLEU | | |
|---|---|---|---|
| | Baseline | Xu's | Our method |
| NIST2002 | 17.11 | 17.22 | 17.37 |
| NIST2003 | 18.84 | 18.52 | 19.03 |
| NIST2004 | 20.35 | 20.07 | 20.60 |
| NIST2005 | 17.82 | 17.12 | 17.10 |
| NIST2006 | 18.22 | 18.59 | 18.58 |
| NIST2008 | 16.86 | 16.72 | 16.92 |

Table 2: Experiment result of different models

Compared with baseline system, (Xu et al., 2005)'s method obtained improvements on test set NIST2002 and NIST2006 and we obtained improvements on all test sets except NIST2005. On NIST2005 and NIST2006, we achieved almost the

[1] http://www.statmt.org/moses/

[2] We remove sentences with length ratio greater than or equal to 9.

| | | |
|---|---|---|
| 1 - Segmentation instance | baseline | 九 〇 年 ， 他 联合 一些 外国 朋友 打电话 给 世界 八个 国家 重要 元首 ， 提醒 他们 环保 的 重要 ， 当作 给 他们 的 礼物 。<br>in 1990 , he got together with some foreign friends to call the leaders of all major countries to remind them of the importance of environmental protection , treating this as a gift to the leaders . |
| | Xu's | 九 〇 年 ， 他<br>in 1990 , he<br>联合 一些 外国 朋友 打电话 给<br>got together with some foreign friends to call<br>世界 八个 国家 重要 元首 ， 提醒<br>the leaders of all major countries to remind<br>他们 环保 的 重要 ， 当作 给 他们 的 礼物 。<br>them of the importance of environmental protection , treating this as a gift to the leaders . |
| | Our method | 九 〇 年 ， 他 联合 一些 外国 朋友<br>in 1990 , he got together with some foreign friends<br>打电话 给 世界 八个 国家 重要 元首 ， 提醒 他们<br>to call the leaders of all major countries to remind them<br>环保 的 重要 ， 当作 给 他们 的 礼物 。<br>of the importance of environmental protection , treating this as a gift to the leaders . |
| 2 - Translation instance | input | （ 法新社 华盛顿 六日 电 ） 美国 总统 布希 今天 用 电话 恭贺 国家 航空 暨 太空 总署 （ nasa ） 官员 的 火星 任务 ， 并 赞誉 他们 的 工作 弥补 哥伦比亚 号 太空 梭 灾难 的 伤痕 。 |
| | baseline | agence france presse washington ( 6 ) the us president george bush &apos;s wish for the use of the telephone national aeronautics and space agency ( nasa ) officials to the moon , and as they work for the space shuttle columbia disaster. |
| | Xu's | . ( courtesy of agence france presse washington 6 ) us president george bush today to phone congratulated national aeronautics and nasa ( nasa ) officials on mars mission praised their work , and the space shuttle columbia compensate the wounds of the disaster. |
| | Our method | on the washington 6. ( courtesy of agence france presse ) us president george bush today the telephone congratulated national aeronautics and nasa ( nasa ) officials on mars mission , and praised their work space shuttle columbia to make up for the scars of the disaster. |

Table 3: Segmentation and Translating Examples

same BLEU scores as (Xu et al., 2005)'s method (the difference is not significant). On the other test sets, we achieved considerable improvement compared with (Xu et al., 2005)'s method (0.15, 0.51, 0.53, 0.20 BLEU scores respectively). For test set NIST2005, both (Xu et al., 2005)'s method and our method had decreased. The sentences in NIST2005 were relatively longer, and the test sentences weren't processed in both decoding stage. These may cause decrease of translation quality.

We further compare the model size for each method. Compared with baseline system, (Xu et al., 2005)'s method and our method had almost the same model size, which had a 13% smaller translation model (phrase translation table) and a 11% smaller reordering model (phrase reordering table). That's because long sentences were split into short sentences, phrase pairs and reordering pairs on the original segmentation positions were not counted. This shows that better result and smaller model size can be achieved with sentence segmentation model.

Table 3 shows some segmentation and translation examples, wherein No.1 is a segmentation instance, No.2 is a translation instance.

## 5 Conclusion

This paper proposes a segmentation method integrating with boundary word probability in training stage for long sentences in statistical machine translation. This method can obtain the boundary words set and their probabilities automatically and give effective guidance for sentence segmentation. Using this method, we can get more meaningful sub sentences in semantics instead of arbitrary segments compared with (Xu et al., 2005)'s method. Besides, we add another factor to balance the length of segmented sub sentences, which will consider the information from both source and target languages. The sentence segmentation model proposed in this paper is independent of languages and statistical translation system, which can be used in either phrase based or syntax based translation systems. Experiments showed that this method could improve translation quality up to 0.5 BLEU scores compared with baseline system.

With the training corpus increases constantly, there would be more and more long sentences. How to use the information in long sentences effectively to improve the translation quality is one of the important problems to be solved. In this paper, we propose a method to improve sentence segmentation in the training process. In the future, we will work on improving long sentence segmentation in decoding stage and combining the translations of segmented sub sentences.

## References

David Chiang. 2007. Hierarchical phrase-based translation. *computational linguistics*, 33(2):201–228.

Takao Doi and Eiichiro Sumita. 2003. Input sentence splitting and translating. In *Proc. of Workshop on Building and Using Parallel Texts, HLT-NAACL 2003*, pages 104–110.

Osamu Furuse, Setsuo Yamada, and Kazuhide Yamamoto. 1998. Splitting long or ill-formed input for robust spoken-language translation. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 421–427. Association for Computational Linguistics.

Yanqing He, Jiajun Zhang, Maoxi Li, Licheng Fang, Yufeng Chen, Yu Zhou, and Chengqing Zong. 2008. The casia statistical machine translation system for iwslt 2008. In *IWSLT*, pages 85–91. Citeseer.

Yeun-Bae Kim and Terumasa Ehara. 1994. A method for partitioning of long japanese sentences with subject resolution in j. *E machine translation*, pages 467–473.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180. Association for Computational Linguistics.

Maoxi Li, Jiajun Zhang, Yu Zhou, and Chengqing Zong. 2009. The casia statistical machine translation system for iwslt 2009. In *IWSLT*, pages 83–90.

Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 609–616. Association for Computational Linguistics.

Qun Liu. 2003. Survey on statistical machine translation [j]. *Journal of Chinese Information Processing*.

Biping Meng, Shujian Huang, Xinyu Dai, and Jiajun Chen. 2009. Segmenting long sentence pairs for statistical machine translation. In *Asian Language Processing, 2009. IALP'09. International Conference on*, pages 53–58. IEEE.

Francisco Nevado, Francisco Casacuberta, and Enrique Vidal. 2003. Parallel corpora segmentation using anchor words. In *Proceedings of the 7th International EAMT workshop on MT and other Language Technology Tools, Improving MT through other Language Technology Tools: Resources and Tools for Building MT*, pages 33–40. Association for Computational Linguistics.

Katsuhito Sudoh, Kevin Duh, Hajime Tsukada, Tsutomu Hirao, and Masaaki Nagata. 2010. Divide and translate: improving long distance reordering in statistical machine translation. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 418–427. Association for Computational Linguistics.

Tong Xiao, Jingbo Zhu, Hao Zhang, and Qiang Li. 2012. Niutrans: an open source toolkit for phrase-based and syntax-based machine translation. In *Proceedings of the ACL 2012 System Demonstrations*, pages 19–24. Association for Computational Linguistics.

Deyi Xiong, Min Zhang, and Haizhou Li. 2010. Learning translation boundaries for phrase-based decoding. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 136–144. Association for Computational Linguistics.

Jia Xu, Richard Zens, and Hermann Ney. 2005. Sentence segmentation using ibm word alignment model 1. In *Proceedings of EAMT 2005 (10th Annual Conference of the European Association for Machine Translation)*, pages 280–287.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 523–530. Association for Computational Linguistics.

Hao Zhang, Daniel Gildea, and David Chiang. 2008. Extracting synchronous grammar rules from word-level alignments in linear time. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 1081–1088. Association for Computational Linguistics.