Language Post Positioned Characteristic Based Chinese-Vietnamese Statistical Machine Translation Method

Jianyalin He[1,2],Zhengtao Yu[1,2],Changtao Lv[1,2], Hua Lai[1,2] , Shengxiang Gao[1,2] ,Yang Zhang[1,2]
[1](School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)
[2](Key Laboratory of Intelligent Information Processing, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)
Correspongding author:Zhengtao Yu, Email:ztyu@hotmail.com

***Abstract*—According to the typical characteristics of Vietnamese modifiers post positioned, a Chinese Vietnamese statistical machine translation method that fuses language post positioned characteristic function was proposed. In this method, firstly we analyzed the grammar differences between Chinese and Vietnamese, and extracted the difference of attribute position, adverbial position and qualifier order. Secondly we defined ordering block based on those difference, and added mapping method of ordering block to decoding algorithm in phrase based statistical machine translation model, then recorded N-best candidate translations and its score P produced by decoding. Thirdly reordered ordering block with language post positioned characteristic reordering algorithm and estimated score D by unconditional maximum likelihood probability distribution. Finally we chose the best translation based on score P and score D. We used lexicalized reordering model based phrase statistical machine translation as contrast experiment. The results of experiment show that our method effectively improves the quality of translation.**

Keywords-statistical machine translation; Chinese; Vietn amese; language post positioned

I. INTRODUCTION

Vietnam is one of the important neighbors in southwest China. The study of Chinese-Vietnamese statistical machine translation plays an important role in supporting Chinese and Vietnamese bilingual comprehension, public opinion analysis, information retrieval and cultural exchange. At present, the translation of Chinese -Vietnamese statistical machine is mainly focused on the construction of Chinese-Vietnamese bilingual resource library, the study of Chinese-Vietnamese word alignment, and translation research is still in its infancy. In the Chinese-Vietnamese word alignment, it significantly improved the accuracy of word alignment after incorporating the Vietnamese attributive post positioned into the log-linear model[1] . In the Vietnamese translation, the US Air Force during the Vietnam War funded the development of the first English-Vietnamese translation system LOGOS I, the system contains only 130 words and has the initial translation function; Yun Zhou et al.[2] developed the Chinese-Vietnamese machine translation system Han Viet 0.1 using the rule-based syntactic transformation method. The system can only convert local subtrees and can not convert the global subtree, and its rules can not cover all the linguistic phenomena. Phuoc Tran et al.[3] and Danhui Yan et al.[4] proposed the SVBUT model and the PVBUT model based on the Chinese-Vietnamese meaning relation, which mainly solves the problem of translation of unknown words. Google translation also integrates Chinese-Vietnamese translation system, but its translation quality is far from meeting the demand. The above-mentioned Chinese Vietnamese machine translation method does not consider the influence of the linguistic features of Chinese and Vietnamese on the translation effect, and the differences in different languages have a significant effect on the performance of the machine translation system. Therefore, how to integrate the language difference between Chinese and Vietnamese language into the machine translation system is a question worthy of study.

In terms of grammatical differences, Vietnamese and Chinese have both similarities and differences in grammar. The same point are the consistency of the main language of the two languages, they are SVO (the subject predicate object). For example, the Chinese is "我听音乐(I listen to music)", as Vietnamese is "Tôi (我) (me) nghe (听) (listen) nhạc (音乐) (music)"; The most striking difference is that the position of modifiers(attributive and adverbial) in the Vietnamese language is in the mirror-image relationship with the Chinese. That is, the adjectives in Vietnamese are located after their modified nouns, the adverbs are located after their modified adjectives and verbs. And the Chinese is just the opposite. For example, the Chinese is "她是一个漂 亮 的 女 孩 (she is the most beautiful girl)", the Vietnamese is "Cô là một (她是) (she is) cô gái (女孩） (girl) xinh đẹp (美丽的) (beautiful) nhất (最)(the most)"[4][5][6][7][8]. After the above analysis can be seen, Vietnamese and Chinese in the word order is significantly different, and these differences have a certain law: The position of modifier and modifier in Vietnamese language is opposite to that in Chinese; Modifiers and modifiers appear consecutively. By analyzing and summarizing these differences, some ordering rules can be constructed; Combine these rules into the phrase-based translation model[9][10][11][12] to explore the impact on the language characteristics of Chinese and Vietnamese on the performance of translation systems.

II. THE VIETNAMESE LANGUAGE POST POSITIONED CHARACTERISTIC

The difference between the language of the Vietnamese and the Chinese modifier (attributive, adverbial) and the central language is the most obvious difference between the two languages. For the problem of the position of the modifier, the attributive of the nouns in the Chinese language are always preceded by the noun, and the language in the Vietnamese language is usually located after the noun. The order for descriptive multi-tiered adverbs in Chinese is: 1) Predicate phrases; 2) Verb

978-1-5386-1981-0/17/$31.00 ©2017 IEEE

(phrase) \ prepositional phrase; 3) Adjective phrases and other descriptive phrases; 4) The adjective and descriptive noun which don't use. The relationship between the Structure of Descriptive Multilanguage Attributes and the Mirror Image in Chinese. The order for descriptive attributive in Chinese is:1-2-3-4 central language; The order for Vietnamese is: central language 1-2-3-4.Table Ⅰ shows an example of the structural characteristics of the Vietnamese-Chinese characters.

TABLE I VIETNAMESE CHINESE WORD STRUCTURE FEATURES

Chinese word order	Vietnamese word order
她是我见过的最美丽的女孩 (She is the most beautiful girl I've ever seen)	Cô là（她是）[she is]cô gái（女孩）[girl]đẹp（美丽的）Nhất[beautiful]（最）[the most]mà tôi từng thấy（我见过的）[I've seen]
孩子们很喜欢那种上面有字的白色的心状糖果 (The children liked the white heart candy with the words)	Trẻ con（孩子们）[children]rất（很）[like]thích(喜欢)[liked]loại（那种）kẹo（心状糖果）[the heart candy]màu trắng（白色 ）[white] ben trên có chữ （上面有字）[the words]

According to the linguistic characteristics of Vietnamese, the grammatical features of Vietnamese are summarized as follows:

1)The adverbs which modified verbs are postponed. That is the adverbs of the modified verbs are located behind the verb;

2)The adverbs which modified adjective are postponed. That is the adverbs of the modified adjectives are located behind the adjectives;

3)The adverbs which modified nouns are postponed. That is the adjective of the modified noun (or noun phrase) is behind the noun (or the noun phrase);

4)Multiple adjectives modified the same noun, adjective reverse order is postponed. That is, multiple adjectives co-modify a noun (or noun phrase), the order for adjectives in reverse orders (relative to the order in Chinese) and arranged behind the modified noun.

From the above four characteristics can be seen two rules: In Vietnamese modifiers are located after the modified words; Modifiers and modifiers are continuous in word order. And also, the latter law exists on Chinese, and it can be easily found in these laws from Table Ⅰ.

Based on the above rules, we consider constructing the function as a feature. At first, three kinds of modulation blocks are defined for Chinese sentences:

1) Verb chronological block: adverb + verb;

2) Adjective chronological block: adverb + adjective;

3) Nouns chronological block: adjective (adjective chronological block) + nouns (noun phrases).

In Vietnamese, the positions of adverbs and verbs, adverbs and adjectives, adjectives and nouns are exactly the reversal as those in Chinese. Therefore, the above three kinds of modulation in the Vietnamese language in the form of the following circumstances:

1) Verb chronological block: verb + adverb;

2) Adjective chronological block: adjective + adverb;

3) Nouns chronological block: nouns (noun phrases) + adjective (adjective chronological block).

According to the three modulation blocks defined above, three sorting rules can be defined to guide the decoding process.

III. THE RECOGNITION METHODS OF LANGUAGE FEATURE ORDERING BLOCK

To use the above rules, we must do word segmentation and POS tagging on the Chinese sentence, but only need to word segmentation on the Vietnamese. In the decoding process, the Chinese POS tagging and ordering block structure can be mapped to Vietnamese. The focus on the Chinese POS tagging is to mark the nouns, adjectives, verbs, adverbs, and the use of Peking University's POS tagging[13]. And the Peking University's POS tagging focus on noun, adjective, verb and adverb. As shown in Table Ⅱ below:

TABLE II CORRESPONDENCE BETWEEN PPOS TAGGING IN CHINESE AND VIETNAMESE

POS	Chinese Mark
noun	n(Common nouns), nr(name), ns(Placename), nt(Institutional team), nz(others), r(pronoun)
adjective	a(adjective)
adverb	d(adverb)
verb	v(verb)

The pronunciation sequence of the Chinese sentence f is: $f = f_{1(m)} = fT_1......fT_i......fT_m$. The segmentation sequence of Vietnamese sentence e is: $e = e_{1(n)} = eT_1......eT_i......eT_n$,an m and n respectively represent the number of word of the sentence f and e. For the Chinese sentences which have been marked by POS, we have to carry out three kinds of ordering block (noun, verb, adjective). The recognition algorithm is described below:

1) If $W_{POS}(i) \in [n, nr, ns, nt, nz, r] \cup \{v\} \cup \{a\}$ then 2) will be carried out, otherwise 1) will be repeated and $W_{POS}(i+1)$ will be judged;

2) If $W_{POS}(i) \in \{n, nr, ns, nz, r\}$ the loop determines whether $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ is an adjective(a). When $W(i-k)(i-k>0, k=1,2,3,...)$ is not adjective(a), the loop will be stopped. And it will contain all the words between $W(i-k)(i-k>0, k=1,2,3,...)$ and $W(i)$ as an ordering block. And the modulation block is represented as a noun(n) otherwise, 3) is executed.

3) If $W_{POS}(i) = v$, the loop determines whether $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ is an adverb (d). When $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ is not adverb (d), the loop will be stopped. And it will contain all the words between $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ and $W(i)$ as an ordering block. And the ordering block is represented as a verb(v); otherwise 4)is executed.

4) If $W_{POS}(i) = a$,the loop determines whether $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ is an adverb (d). When $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$is not adverb$(d)$, the loop will be stopped. And it will contain all the words between $W_{POS}(i-k)(i-k>0, k=1,2,3,...)$ and $W(i)$ as an ordering block. And the ordering block is represented as an adjective(a); otherwise 1) will be repeated and $W_{POS}(i+1)$ will be judged.

5) For other the word of POS (except nouns, verbs, adjectives, adverbs other than words), we do not deal with.

The above rules can be expressed as the following functions:

$$C_{i(i=n,v,a)}\begin{cases} C_n & if\ W(i)=(n,nr,ns,nt,nz,r)\cup W(i-k)\neq a \\ C_v & if\ W(i)=(v)\cup W(i-k)\neq d \\ C_a & if\ W(i)=(a)\cup W(i-k)\neq d \end{cases} \quad (3)$$

Where C_i is an ordering block, C_n is an ordering block of noun, C_v is an ordering block of verb, C_a is an ordering block of adjective.

According to the result of block recognition, we can adjust the sequence of intra phrase reordering for the decoded N-best candidate translation.

IV. Chinese - Vietnamese Statistical Machine Translation with Language Post Positioned Characteristic

4.1 Phrase-Based Statistical Machine Translation

The translation framework of this paper is phrase-based statistical machine translation[14][15][16][17] . The phrase-based statistical machine translation can be described as: The source language f is divided into m phrases (consecutive word sequences), where $f = fT_1,...,fT_i,...,fT_m$; And then translates each phrase fT_i into the target language phrase eT_i; And then generating the target language sentence $e = eT_1,...,eT_i,...,eT_n$, and it will find the highest probability of translation in all possible target translations. The phrase-based statistical machine translation model can be described as:

$$\begin{aligned} e_{best} &= \arg\max_e \prod_{i=1}^{m} p\left(\overline{f_i}\middle|\overline{e_i}\right) \\ &\cdot \prod_{i=1}^{|e|} p_{LM}\left(e_i\middle|e_1...e_{i-1}\right) \\ &\cdot D(i) \end{aligned} \quad (4)$$

Where $P\left(\overline{f_i}\middle|\overline{e_i}\right)$ is the translation model, $P_{LM}\left(e_i\middle|e_1...e_{i-1}\right)$ is language model, $D(i)$ is Order model.

4.1 Decoding

Using CKY Decoder to Decode Chinese-Vietnamese Statistical Machine Translation. In order to apply the ordering rules of the ordering block, it is necessary to map the block C_i in the source language sentence to the Vietnamese candidate translation. The mapping rules of the ordering block C_i are shown in Table III below:

TABLE III Mapping rules for the ordering block C_i from the source language to the target language

Ordering block	Chinese $C_{ci(i=n,a,v)}$	Vietnamese $C_{vi(i=n,a,v)}$
C_v	慢慢地(slowly)/d/Cv 走(go)/v/Cv	từ từ/d/Cv đi/v/Cv
C_a	最(the most)/d/Ca 美丽的(beautiful)/a/Ca	nhất/d/Ca đẹp/a/Ca
C_n	美丽的(beautiful)/a/Cn 女孩(girl)/n/Cn	đẹp/a/Cncô gái/n/Cn

The CKY decoder is improved according to the mapping rules of the ordering block C_i. When the decoder searches for candidate translations of the source language words, the word POS and the ordering block C_i of the source language are mapped to the candidate translation. The input of the modified CKY decoder is the source language sentence $f = fT_1,...,fT_i,...,fT_m$ and its ordering block structure C_{ci} after the word segmentation, POS tag and ordering block recognition; The output is a N-best candidate translation table with the POS tag and the ordering block tag C_{vi} ; And record the corresponding sentence score P, according to the score P in reverse order, get the candidate sentence P sort value $index_P$. The improved CKY decoder algorithm pseudo-code shown in Figure 1.

```
Input: source string f=f1…fi,and Ci
place empty hypothesis into stack 0
for all stacks 0...n-1 do
    for all hypotheses in stack do/
        for all translation options do
            if applicable then
                create new hypothesis
                if hypothesis belongs to Ci then
                    tag POS to hypothesis
                end if
                place hypothesis in stack
                recombine with exiting hypothesis if possible
                prune stack if too big
            end if
end for
end for
end for
Output: N-best candidate translations
```

Figure 1. Improved CKY Decoder Algorithm Pseudo-Code

The N-best candidate translation for decoding is reordered by the ordering rules of the ordering block. In Vietnamese, all the modifiers are located behind the modified words it modifies(Having multiple modifiers followed by the modifier in close proximity to the modified word). So we have only one sort of rules for the ordering block, that is, descend. The form is as follows equation (5):

$$\begin{aligned} C_{vi(i=n,v,a)} &\rightarrow X_1X_2...X_n, \\ &X_nX_{n-1}...X_1 \end{aligned} \quad (5)$$

When it is ordering ,at first, it should check whether the phrase inside the Vietnamese chunks is relative to the descending of the phrase inside the Chinese tuple. If it is, the ordering will stop, the direct adjustment of the ordering block score. If not, adjust the order for the Vietnamese chunks to the phrase descend on the context of the Chinese tuple. And calculate the order for the adjusted order block. The internal ordering algorithm of ordering block is shown in Figure 2.

```
Input: N-best candidate translations stack
for all candidate translation do
    for all reordering Ci do
        if descend then
                calculate score
                place in stack else
                descend order
        calculate score
        place in stack
    end if
    end for
end for
Output: N-best candidate translations
```

Figure 2. The Internal Ordering Algorithm of Ordering Block

After sorting in accordance with the above rules, you need to re-estimate the N-best candidate translation score. Since the order is only for the internal order of the Vietnamese chunks C_{vi}, the N-best candidate's translation can be estimated by counting the frequency of the appearances in the training corpus by counting the Chinese-Vietnamese chunks (C_{ci}, C_{vi}) N-best candidate translation score; In order to avoid data sparseness, the data smoothing is performed using an unconditional maximum likelihood distribution of factor σ . The score for the N-best candidate translation is given by the following equation (6):

$$P_o\left(descend \mid \overline{f}, \overline{e}\right) = \frac{\sigma p(descend) + count\left(descend, \overline{f}, \overline{e}\right)}{\sigma + \sum_o count\left(o, \overline{f}, \overline{e}\right)} \quad (6)$$

Where, $descend$ represents the ordering rule (as equation(5)), o represents all sequential combinations of the Vietnamese tone block internal phrases. The definition of $P(descend)$ is as follows the equation (7):

$$P_o(descend) = \frac{\sum_e count\ (descend, \overline{f}, \overline{e})}{\sum_o \sum_e count\left(o, \overline{f}, \overline{e}\right)} \quad (7)$$

According to the equation (6), (7) can be calculated with the order of the order of N-best candidate translation after the score $P_o\left(descend \mid \overline{f}, \overline{e}\right)$; And then get the sort value $index_d$ of each candidate sentence according to $P_o\left(descend \mid \overline{f}, \overline{e}\right)$; And then, the score of the i-th sentence in the N-best candidate translation is corrected as $Score(i)$:

$$Score\ (i) = \frac{index_P + index_d}{2} \quad (8)$$

N-best candidate translations are reordered with ne w scores. And finally, the smallest value of one or N sentences output, as the final translation results.

V. EXPERIMENTS

5.1 Experimental Data

As a result of the lack of authoritative Chinese - Vietnamese bilingual sentences on the parallel corpus, the experimental use of Chinese and Vietnamese parallel corpus are from the laboratory, corpus is crawling the Chinese - Vietnamese chapter aligned corpus from Internet. And then through the computer sentence alignment and artificial proof method to obtain. At present, the collection of Chinese and Vietnamese parallel sentence is about 25,000. The details of the experimental data are shown in Table III below.

TABLE VII THE CORPUS INFORMATION USED IN THE EXPERIMENT

		Chinese	Vietnamese
Training set	Sentence	25000	
	Word	275659	275659
Development set	Sentence	400	
	Word	47962	47962
Language model training set	Sentence	2000	
	Word	22763	22763
Test set	Sentence	100	
	Word	1142	1142

5.2 Experimental Setting and Experimental Results

The experiment is based on the phrase model system in Niutrans open source system[18] as a reference prototype. And we will use the tools developed by our labs to carry out lexical analysis of Vietnamese. And we use GIZA++ to get the Chinese-Vietnamese word alignment results. The ternary language model is trained by the Niutrans machine translation platform. We use the BLEU evaluation criteria as an evaluation tool. In the experiment, we designed three groups of contrast experiments to test the effect of ordering sequence regulation on Chinese-Vietnamese machine translation, it is respectively: fusion of language post positioned characteristic Features Chinese - Vietnamese phrase machine translation method(Phase-based SMT + reorder rules), Fusion of lexical ordering model (MSD)[19][20][21] Chinese-Vietnamese phrase machine translation method (Phase-based SMT + MSD), fusion of Language Postposition and Lexical of ordering model (MSD) Chinese Vietnamese phrase machine translation (Phase-based SMT + MSD + reorder rules). The experimental results are shown in Table IV.

TABLE V CHINESE-VIETNAMESE MACHINE TRANSLATION SYSTEM COMPARATIVE EXPERIMENT

Method	BLEU(%)
Phase-based SMT + reorder rules	18.76
Phase-based SMT + MSD	21.03
Phase-based SMT + MSD + reorder rules	24.15

It can be seen from Table IV that the translation result is relatively poor when the language translation of Chinese-Vietnamese phrases is used only by language post positioned. The reason may be that the rule-based ordering model does not cover all language phenomena and can not be ordered for language that can not be covered; The translation effect of Chinese - Vietnamese phrase is only improved by lexicalization, but the effect of language feature of translation effect can not be solved; The translation effect is improved significantly when the lexical ordering model and the language post positioned feature are used together for the translation of Chinese-Vietnamese phrases. The experimental results show that the Chinese-Vietnamese phrase statistical machine translation method, which integrates language post positioned feature, can effectively improve the accuracy of translation.

VI. CONCLUSIONS

For the characteristics of the typical language of Vietnamese language, this paper proposes a Chinese-Vietnamese statistical machine translation method which combines the post positioned feature function of language. This method is based on the Chinese-Vietnamese phrase model, and adds the Chinese-Vietnamese ordering block mapping method of the decoding algorithm. And the language post positioned feature is used to optimize the sequence of candidate translation tables generated by decoding. Experiments show the effectiveness of the proposed method. The fusion language post positioned feature function can improve the accuracy of translation. The further study focuses on the construction of Chinese-Vietnamese bilingual translation models and the optimization of performance.

ACKNOWLEDGMENT

This work was supported by National Nature Science Foundation (Grant Nos.61472168, 61672271, 61732005), and Science and Technology Innovation Talents Fund Project of Ministry of Science and Technology(Grant No.2014HE001), Innovation Team Project of Yunnan Province(Grant No.2014HC012). The corresponding author is Zhengtao Yu. His email is ztyu@hotmail.com.

REFERENCES

[1] Yuanyuan Mo, Jianyi Guo, Zhengtao Yu, etc..A bilingual word alignment algorithm of Vietnamese-Chinese based on feature constraint[J]. International Journal of Machine Learning & Cybernetics, 2015, (6):538-543

[2] Yun Zhou..Chinese - Vietnamese Machine Translation Experiment System[D]. Henan Luoyang : People 's Liberation Army Foreign Language Institute, 2006.

[3] Phuoc Tran, Dien Dinh, and Linh Tran.Resolving Named Entity Unknown Word in Chinese-Vietnamese Machine Translation[J].Advances in Intelligent Systems & Computing, 2014, (245):273-284

[4] Danhui Yan,,Yude Bi.Research on the Recognition of Vietnamese Naming Entity Based on Rule[J].Journal of Chinese Information Processing , 2014, 28(5):198-205

[5] Shichunron Wu.On the Influence of Chinese on Vietnamese[J]. Journal of University of Jinan(Social Science Edition), 2001, 11(5):56-57

[6] Shihe Wu. Compared Study of the Syntactic Word of Vietnamese and Chinese[J]. Journal of Yunnan Normal University:Teaching and Research of Chinese as A Foreign Language , 2005, 3(6):65-68

[7] Shihongfeng He.Compared Study of "Attributive" in Chinese and Vietnamese[D]. Wuhan, Hubei : Central China Normal University, 2006.

[8] Qingchuan Fan. Compared Analysis of the Word Order for Chinese and Vietnamese[J]. Journal of Yunnan Normal University:Teaching and Research of Chinese as A Foreign Language, 2007, 5(6):79-85

[9] Tong Xiao, Hao zhang, Qiang Li, etc.The niutrans machine translation system for cwmt2011[C]//Proceeding of the 6th China Workshop on Machine Translation, Xiamen, China, August 2011. CWMT.

[10] Deyi Xiong, Qun Liu, Shouxun Lin.2006. Maximum entropy based phrase reordering model for statistical machine translation[C]// Proceeding of COLING/ACL 2006, Sydney, Australia, 2006, pages 521-528.

[11] Daniel Marcu and William Wong.A phrase-based, jointed probability model for statistical machine translation[C]// Proceedings of EMNLP 2002, Philadelphia, PA, USA, 2002, pages 133-139.

[12] Bo Xu, Xiao Dong Shi, Qun Liu, et al.Current Statistical Machine Translation Research in China[J]. Journal of Chinese Information Processing, 2006, 20(5):1-9

[13] Shiwen Yu, Huiming Duan, Xuefeng Zhu,.Basic Code of Modern Chinese Corpus in Peking University[J]. ournal of Chinese Information Processing, 2002, 16(5):49-64

[14] Philipp Koehn, Franz Joseph Och, and Daniel Marcu.Statistical Phrase-Based Translation[C]// Proceedings of HLT-NAACL 2003, Alberta, Canada, 2003, pages 127-133.

[15] Franz J.OchStatistical machine translation: from single word models to alignment templates[J]. Rwth Aachen, 2002, 10(2):65-70

[16] Yaster Al-Onaizan, Kishore Papineni.Distortion Models for Statistical Machine Translation[C]// Proceedings of COLING/ACL, Sydney, Australia, 2006:529-536.

[17] Chris Quirk, Arul Menezes.Do we need phrases? Challenging the conventional wisdom in statistical machine translation[C]// Proceeding of NAACL, New York, USA, 2006:677 - 678.

[18] Tong Xiao, Jingbo Zhu, Hao Zhang, etc.NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation[C]// Proceeding of ACL 2012, Jeju Island, Korea, 2012, demonstration session.

[19] Christoph Tillmann.A unigram orientation model for statistical machine translation[C]// Proceedings of HLT-NAACL 2004: Short Papers. Association for Computational Linguistics, 2004:101-104.

[20] Koehn P, Axelrod A, Mayne A B, et al.2005. Edinburgh system description for the 2005 NIST MT evaluation[C]// Proceedings of Machine Translation Evaluation Workshop.

[21] insong Su, Yang Liu, Yajuan Lu, etc.Learning Lexicalized Reordering Models from Reordering Graphs[C]// Proceeding of Meeting of the Association for Computational Linguistics. 2010:12-16.