Local Phrase Reordering Model for Chinese-English Patent Machine Translation

Xiaodie Liu
Institute of Chinese Information Processing, Beijing Normal University, Beijing 100875, China
liuxiaodie2009@hotmail.com

Yun Zhu
Institute of Chinese Information Processing, Beijing Normal University, Beijing 100875, China
zhuyun@bnu.edu.com

Yaohong Jin
Institute of Chinese Information Processing, Beijing Normal University, Beijing 100875, China
jinyaohong@bnu.edu.cn

Abstract: We focused on when and how to reorder the Chinese NPs with two, three translation units for reordering in Chinese-English Machine Translation. By analyzing the features of translation units of the Chinese NPs, we built some formalized rules to recognize the boundaries of translation units using the boundary words to recognize what to reorder. By comparing the orders of Chinese and English NPs, we developed a strategy a local phrase reordering model on how to reorder the translation units. At last, we used a rule-based MT system to test our work, and the experimental results showed that our rule-based method and strategy were very efficient.

1 Introduction

Over the past decade, machine translation (MT) has seen many exciting developments, but phrase-based models, syntax-based models, dependency-based models cannot deal with the reordering at the sentence level very well, particularly the reordering in a hierarchical structure in which the reordering of Chinese NPs is dependent on. The Hierarchical Semantic-Category-Tree (HSCT) Model[1] used semantic features to handle the reordering at sentence level with a rule-based method. The system based on HSCT model could partition a sentence into a predicate and noun phases (NPs for short) and reorder them in sentence level, but this system did not reorder the inner orders in a NP. This paper aims to solve a type of the problem: reorder the inner components in a NP.

In Chinese-English MT, the structural difference in NPs between Chinese to English is a difficult problem, such as the different positions for the head of the NPs: the head of a Chinese NP is in the front; however that of an English NP is in the end. The language for patent usually was inclined to express complicated thought in long and complex words. Here is an example wherein the number of the translation units in a complicated Chinese NP happens to be three.

Example:

Chinese NP:所述图像传感器于红色、蓝色和绿色色彩通道的光谱灵敏度

English in Chinese order: the image sensor device in red, blue, and green color channels de the spectral sensitivities

Reference: the spectral sensitivities of the image sensor device in red, blue, and green color channels

Google: an image sensor means in said red, blue and green spectral sensitivity of color channel.

The Chinese NP “所述图像传感器于红色、蓝色和绿色色彩通道的光谱灵敏度”consists of a Chinese character “的” and a function word “于” and three translation units: A“所述图像传感器”、B“于红色、蓝色和绿色色彩通道” and C“光谱灵敏度”, and the right English order of translation units was “C A B” according to the reference. However, the boundaries of “B and C” and the order of “A, B and C” in English of Google were wrong.

How to recognize the boundaries of the translation units “A, B, C” and how to reorder them were the main problems that to be solved in this paper.

The remainder of this paper is organized as follows. In Section 2, we at first the examined the features of translation units and sum boundary words for recognizing them, then analyzed differences in the orders of complicated Chinese-English NPs and summed a strategy on how to reorder the translation units when translate the complicated Chinese NPs to English. In Section 3, we used a rule-based method and designed a algorithm to recognize the translation units to solve what to reorder and how to reorder. In Section 4, we discussed the experiment results. Section 5 is the related work. Finally, a

conclusion is given and the further work is expected in Section 6.

2 Analyses of the Chinese NPs

In this section, we intent to develop a strategy on reorder Chinese NP based on the translation units.

2.1 Translation units

This part focused on examining the features of the translation units and tried to sum some laws on the boundaries for recognizing the translation units.

A translation unit was a word or word-group of expressing the same meaning both in Chinese and English. Translation units were the equivalent in Chinese and English NPs, and it is certain that this study can only be based on contrasting Chinese and English NPs. The complicated Chinese NP consists of NP or prepositional phrases (PP for short), locative phrase (LP for short) etc. The translation unit may be as follows:

(1) NP, which was a smallest unit for reordering in Chinese-English machine translation;

(2)PP with a "preposition" and (1)

(3)LP with a "localizer" and (1)

(4)PP with a "preposition", (1) and a "localizer"

(5) PP with a "preposition", some (1) and a "localizer".

2.1.1 Outer Characteristics

In English, prepositions with flexible meanings are frequently used. They function as bonding agents in NPs. In English NPs, the translation units can be obtained by the prepositions, such as "the spectral sensitivities of the image sensor device in red, blue, and green color channels", the prepositions "of" and "in" can partition the NP into three translation units.

In Chinese NPs, the translation units can be obtained by the prepositions, localizers and the structural particle "的(de)".

- structural particle "的(de)"

Compare and analyze the difference as well as locations between English modifiers and Chinese auxiliary words so that the "的(de)" is especial, which servers two functions. First, "的(de)", marked as "de_1", was a boundary of two translation units and should be excluded from the two translation units, such as the "的" in the phrase "本发明的实施方式", which consist two translation units "本发明"and "实施方式". "的$_2$"; second,"的(de)", marked as "de_2" , is an inner conjunction of a translation unit, such as the "的" in the phrase "简明的装置".

Here we focuses on "de_1" and should distinguish it from "de_2". Based on the studying of the "的(de)",we discovered some language rules from authentic contexts: "de_2" were often behind of adjective, quantifier, verb and pronoun. In view of the cost and practicability, "de_2"can be eliminated more easily.

- prepositions

Preposition is one of Chinese functional words which have a complex function, involving various factors, such as "在(zai),根据(gen ju),作为(zuo wei)"etc. For example,

- localizers

There is a special grammatical method in Chinese language—the words of locality which represent pure directions, such as "上(on),中(in),外(outside)".

- others

Some time nouns, such as "时(when)"and some auxiliary words, such as "而言(er yan)".

Some prepositions and some localizers work hand in hand, such as "当……时", "在……中", "对……而言"etc.

The prepositions and localizers always worked with "de_1", for example, for Chinese NP "一种或多种作为湿润剂的醇类",the preposition "作为" and structural particle "de1" could partition it into three translation units, for Chinese NP "通常运转时的气体压力", the localizer "时"and structural particle "de1" could partition it into two translation units.

From the above analysis we could see structural particle " 的 (de)", prepositions, localizers and others could be outer boundaries words for recognize the translation units.

2.1.2 Inner Characteristics

In Chinese NPs, the NPs may modify the NPs without any prepositions, localizers and structural particle, but not in English NPs.

This section explores some laws at the beginning word and the end word of translation units in Chinese-English translation.

2.1.2.1 Left Characteristics

Chinese and English belong to two different

language families, in most cases, the source language cannot correspond to the target language well. However, the internal components of translation units shared some characters in common as the result of linguistic universality and identity of human thinking form.

By comparing the translation units, we can obtain some features of the beginning of translation units, which can be some implicit boundaries of two translation units.

- demonstrative pronoun

In English NPs, the translation units also can be obtained by the definite articles and demonstrative pronoun. In Chinese NPs, the translation units also can be obtained by demonstrative pronoun. For example, the word “这些(these)” is demonstrative pronoun in translation unit“这些业务(these services)”

- quantitative phrases

In Chinese NPs, the translation units could not obtained by indefinite articles and quantitative phrases. For example, the Chinese phases “一个,一种” and so on correspond to the English words “a, an, one” .

The Chinese words “第一, 第二, 第三, 第四” correspond to the English words “first, second, third, fourth”.

- adjectives

An adjective usually modifies the noun.

The difference of Chinese and English exists in many ways such as morphological structure and word-formation which leads to the different words for a semantic feature between Chinese and English. Some adjectives, past participles, present participles in English were expressed in phase in Chinese.

- degree adverbs

The English adjective “notable” in translation unit “notable features” which are unique for English, thus appropriate version cannot find in Chinese vocabulary and expressed in phases “很显著” in translation unit “很显著的特点”, wherein “很” is a degree adverb. There are many degree adverbs in Chinese, for example " 很(very), 非常(extremely), 十分(very much), 特别(specially), 极(too),(little), 更(more), 较(better),比较(better), 最(best)" etc.

- time adverbs

The past participle “coated” in translation unit “coated tank” which are unique for English, thus appropriate version cannot find in Chinese vocabulary and expressed in phases “已涂覆” in translation unit “已涂覆罐”, wherein “已” is a time adverb. There are many time adverbs in Chinese, for example " 已经(already), 曾经(once), 早已(already), 刚刚(just now), 正(be being),正在(be being),就(be going to),就要(be going to),将(be going to),将要(be going to),曾(once),刚(just now),才(already),在(be being)" etc.

- negative adverbs

The English adjective negative prefixes, such as “negative, opposite, and reverse” etc, corresponds to the negative adverbial words “不(un-),非(un-),没(dis-),没有(dis-),不用（un-）,未（dis-）”etc.

- others

The “resulting” in translation unit “resulting decoded” which are unique for English, thus appropriate version cannot find in Chinese vocabulary and expressed in phases “经解码” in translation unit “经解码语音”.

The past participle “refrigerated” in translation unit “refrigerated substances” which are unique for English, thus appropriate version cannot find in Chinese vocabulary and expressed in phases “被制冷” in translation unit “被制冷物质”, wherein “被” is an auxiliary word. There are many auxiliary words in Chinese, such as “所(suo)” etc.

The English adjective “movable” in translation unit “rotatable structure” which are unique for English, thus appropriate version cannot find in Chinese vocabulary and expressed in phases “可旋转” in translation unit “可旋转结构”.

2.1.2.1 Right Characteristics

- Noun

In general, the end of a translation unit in Chinese patent was a noun.

- number words

The number words consist of the numbers from 0 to 9 and were labels related to the new inventions. Huge numbers of number words could not enter the knowledge base and were generated dynamically as NUM when knowledge

base was loaded.

- letter words

The letter words consist of numbers from 0 to 9 and 26 English letters. As with the number words, Huge numbers of number words could not enter the knowledge base and were generated dynamically as SPN when knowledge base was loaded.

- auxiliary word “等(Deng)”

The auxiliary word “等(Deng)” Displays the enumeration entry.

As can be seen above, although Chinese does not have much overt morphology, it still distinguishes a translation unit from another by lexical devices.

When they were not outer boundary words in one continuous string, we used demonstrative pronoun, degree adverb, quantitative phrase, adjective, time adverbs, negative adverbs and others as the left boundary words to partition the string into two translation units utilizing contextual clues such as right characteristics. See table 1 for examples.

When they were not outer boundary words in one continuous string, we used number words, letter words, auxiliary word “等(Deng)” as the right boundary words to partition the string into two translation units.See table 2 for examples.

Table1.The left boundary words and examples

Left Boundary Words	Examples
demonstrative pronoun	(如权利要求 5)(所述的方法)->(the method) (according to claim 5)
degree adverb	(这项发明)(非常显著地一个特点)->(a very significant feature) of (this invention)
quantitative phrase	(这种丙烯酸树脂)(一个来源)->(one suitable source) of (these acrylic acrylate resins)
adjective	(支撑表面)(可能的总体压力)->(the possible total fluid pressure)of(a support surface)
time adverb	(该容器)(已经清洗的表面)-> (the cleaned surface) of (the containers)
negative adverb	(取向膜) (未拉伸的酰化纤维素膜)-> (unstretched cellulose acylate films) of (Orientation film)
others	(该工件)(被照射区)-> (the irradiated region) of (the work)

Table 2.The right boundary words and examples

Right Boundary Words	Examples
number words	(水箱 300)(底部组件 303) ->(a base assembly 303)(for the tank 300)
letter words	(主体部 400a)(膛)-> (the bore)of (the host section 400a)
等(deng)	(牵引电动机等) (负载装置)->(a load device)(such as the traction motor)

Taken together, the boundary words have two types: 1) The out boundary words ,such as “的(de)” , propositions and locatives; 2) The inner boundary words, which include left boundary words and right boundary words. The left boundary words lie in the first word or phrase in a translation unit and the right boundary words lie in the last word.

2.1 Reordering

Once the translation units had been recognized, a new process called reordering can be applied to them.

This part compared and analyzed the differences in the orders of Chinese-English NPs so as to find out the law and develop a strategy on how to translate the complicated Chinese NPs to English effectively.

According to the boundary words above, we found that there are three types of translation units:

- PPs

PPs with a preposition, such as “根据该形式” NP “根据该形式的模具”, PPs with a preposition and a localizer, such as “在外模具部件 601 和管子 400 外表面之间” in NP “在外模具部件 601 和管子 400 外表面之间的界面”

- LPs

LPs are unique to Chinese, such as “主体上” in NP“主体上的主要压力”,

- NPs

It could be one or more word, such as "/详视图/(a detailed view), /所述/局部/缩放/估算/器/(said local scale estimator)" and could include a "de_2" , such as 大部分常规的交通工具布线系统(most conventional vehicle wiring systems)" .

- NP_Bs

NP_Bs is unique to Chinese. It could be adjectives, pronouns, numerals and quantitative phrases in special position, such as "一些" which could occur in the NP "一些基于环保溶剂的清洗剂" which consist of three translation units "一些", "基于环保溶剂" and "清洗剂".

The Chinese NPs consist of NPs, PPs, LPs and NP_Bs; otherwise, the English NPs consist of NPs and PPs only.

The Chinese and English NPs share some characters in common:1) the NPs consist of PPs and NPs, 2) the PPs could not be head translation units, 3) the NPs must be the head translation units.

However, the position and order between the NPs and PPs in Chinese were different from English as the result of different culture and tradition, especially, the NP_Bs, LPS are unique for Chinese NPs. The primary dissimilarity is that the head translation unit must locate in the end and others translation units such as NPs, PPs and LPs must be modifiers to some extent, otherwise, in English, the beginning translation unit is the head translation unit and followed by other PPs as modifiers.

Based on aligned Chinese-English NP-pairs from 15 Patent documents, this paper compared the differences in orders of Chinese-English NPs and analyzed types of translation units so as to find out the laws in translating Chinese NPs to English.

For Chinese NPs, the PPs and LPs shared same function, so we use PPs for PPs and LPs.

2.1.1 Chinese NPs with two translation units

All combinations and the structure relations about two translation units can be listed in table 2, and we can see that in combinations of "NP1 NP2", the NP1 or NP2 were the NPs and the NP1 must modify NP2; in combinations of "PP NP", the NP must be the NPs and PP must modify the NP.

For the PPs and LPs, the prepositions were before the NPs and the locatives were behind the NPs, but the PPs or LPs could be moved as a whole. In table 2, we can see that there only one reordering way for the Chinese NPs with two translation units and b).

All structure relations between two translation units occur in Chinese, and we can find some examples to illustrate our reduction were right in table 3.

Table3: Chinese and English orders of NPs with two translation units

	Order in Chinese		Order in English
	Combinations	Structure Relations	
a)	NP1 NP2[1]	NP1 NP2	NP2 NP1
b)	PP[2] NP	PP NP	NP PP

Table4: examples of orders of Chinese and English NPs with two translation units

Chinese Orders -> English orders
NP1 NP2-> NP2 NP1 (模制工艺)的(操作参数)->the operating parameters of the molding process
PP NP -> NP PP (在前面实施方案中)(所述的相同类型的聚合材料)-> the same types of polymeric material in the earlier embodiment

2.1.2 Chinese NPs with three translation units

We can list all combinations and the structure relations about three translation units in the Chinese NPs. In table 4, we can see that: a) has three structure relations: (NP1 (NP2 NP3)) indicated that at first NP2 modified NP3, then they acted as a whole to be modified by NP1, ((NP1 NP2) NP3) indicated that at first NP1 modified NP2, then they acted as a whole to modify NP3,and ((NP1) (NP2) NP3) indicated that NP1 and NP2 respectively modify NP3.b) has three structure relations: ((PP NP1) NP2) indicated that at first NP1 modified NP2, then they acted as a whole to be modified by PP, ((PP NP1) NP2) indicated that at first PP modified NP1, then they acted as a whole to modify NP2,and ((PP1)(NP2) NP3) indicated that PP and NP1 respectively modify NP2, and c) has three structure relations:(PP1 (PP2 NP)) indicated that at first PP2 modified NP, then they acted as a whole to be modified by PP1, ((PP1 PP2) NP) indicated that PP1 and PP2 acted as a whole to modify NP, and ((PP1)(PP2) NP)

[1] The number behind NP or PP notes the order they appear in the linear sequence of Chinese NPs. The space among NP and PP is used to split the different translation units, and can stand for "的" in authentic Chinese NPs.

[2]

indicated that PP1 and PP2 respectively modify NP. d) has two structure relations: (NP1 (PP NP2)) indicated that at first PP modified NP2, then they acted as a whole to be modified by NP1, and ((NP1)(PP)NP2) indicated that NP1 and PP respectively modify NP2.e) has one structure relation (NP_B (PP NP)) which indicated at first PP modified NP, then they acted as a whole to be modified by NP_B.

Although, not all the structure relations among three translation units occur in Chinese, we can find some examples to illustrate our reduction were right in table 5.

Table5: Chinese and English orders of NPs with three translation units

	Order in Chinese		Order in English
	Combinations"	Structure Relations	
a)	NP1 NP2 NP3	(NP1(NP2 NP3))	NP3 NP2 NP1
		((NP1 NP2) NP3)	
		((NP1)(NP2) NP3)	
b)	PP NP1 NP2	(PP (NP1 NP2))	NP2 NP1 PP
		((PP NP1) NP2)	
		((PP)(NP1) NP2)	
c)	PP1 PP2 NP	(PP1 (PP2 NP))	NP PP2 PP1
		((PP1)(PP2) NP)	
		((PP1 PP2) NP)	NP PP1 PP2
d)	NP1 PP NP2	(NP1 (PP NP2))	NP2 PP NP1
		((NP1)(PP)NP2)	
e)	NP_B PP NP	(NP_B (PP NP))	NP_B NP PP

Table 6: Examples of orders of Chinese and English NPs with three translation units

Chinese Orders -> English Orders
1 NP1 NP2 NP3 -> NP3 NP2 NP1 1.1((材料)的(捏合度))的(调节范围)->the adjusting range for the kneading degree of the material =>(NP1 NP2) NP3 -> NP3 NP2 NP1 1.2(这些模具)的((一种形式)的(结构))-> the construction of one form of the moulds =>NP1 (NP2 NP3) -> NP3 NP2 NP1
2 PP1 PP2 NP-> NP PP2 PP1 (闸板部件接近于圆筒段时)的(材料间)的(捏合度)->the kneading degree between the material when the gate member is moved close to the cylindrical segment =>PP1 PP2 NP-> NP1 PP2 PP1
3 PP NP1 NP2-> NP2 NP1 PP 3.1 ((在内模具部件上)的(第一密封装置))的(优点)-> the high vapor pressure of the propellant in the MDI =>(PP NP1) NP2-> NP2 NP1 PP 3.2 (调节捏合度时)的((装置)的(操作性))->the operability of the apparatus at the time of adjusting the kneading degree =>PP (NP1 NP2)-> NP2 NP1 PP
4 NP1 PP NP2-> NP2 NP1 PP (上述旋转轴部)的((绕水平轴)的(旋转力))-> the rotation force around the horizontal axis of the rotation axis part =>NP1 PP+ +NP2-> NP2 NP1 PP
5 NP_B PP NP-> NP_B NP PP (一些) ((基于环保溶剂)的(清洗剂))-> some cleaner based on environment-friendly solvent =>NP_B PP NP-> NP_B NP PP

We all know that ambiguity can be caused by multiple and different semantic relations. So, how to reorder seems to be a big problem. However we could conclude some laws about how to reorder the English orders of Chinese NPs with three translation units as follows:

1) The reversed order is the only way for Chinese NPs “NP1 NP2 NP3” which had three modifications and three semantic relations;

2) The first NP and the reversed order for PP is the only way for Chinese NPs “PP PP NP”. For the pattern “((PP1 PP2) NP)->NP PP1 PP2”, it only occurs in a kind of special Chinese NPs, such as “从(from)……”, “到/向(to)……”;

3) The first NP2, second NP1 and third PP is the only way for Chinese NPs “PP NP1 NP2” which had three structure relations and for Chinese NPs “NP1 PP NP2”. For the pattern is (NP1 PP NP2)->NP2 PP NP1, the PP in English has the front boundary, and there is ambiguity, too;

4) The first NP_E, second NP and third PP is the only way for Chinese NPs “NP_B PP NP” in e).

3 Method

In our existing MT system, the Chinese NPs were given, but how to recognize the two or three translation units in the Chinese NPs were unknown. They can be obtained by using boundary words, for an effective boundary word or a combination of boundary words could not be the end or beginning of a base NP in semantic and partition a language string into two translation units.

3.1 Recognitions

The boundary words have two types: 1) the out boundary word ,such as “的(de)” and propositions, 2) the inner boundary words, which include left boundary words and right boundary words. The left boundary words lie in the first word or phrase in a translation unit and the right boundary words lie in the last word.

The Chinese NPs are stored using a tree structure. Thus how to distinguish the “的(de)” into “de_1” and “de_2”, how to recognize the front and rear boundaries and how to combine the words of a unit is important.

Using a rule-base method, we designed some tags and attributes for nodes to recognize the translation units:

1) Tag for Nodes

- MK
  It was a node for “de_1” and means there is a reordering operation.
- MK_Q
  It need to add a new node “de_1” before the left boundary words.
- MK_H
  It need to add a new node “de_1” after the right boundary words.
- L1
  L1s were the front boundaries of translation units.
- L1H
  L1Hs were the rear boundaries of translation units
- NP_Bng%
  NP_Bng% was the beginning position of an Chinese NP.
- NP_End%
  NP_END% was the end position of an Chinese NP.

2) Attributes for Node

- LEVEL
  It was used to put this attribute to “的(de)”,if value=-1, recognize it as “de_1”, if value=0 (default value), recognize it as “de_2”.
- NOT_CHANGE
  It was used to keep a translation unit not moving.

We designed the algorithm as follows:

Step1: distinguishing the “的(de)” into“de_1” and “de_2”

- building 12 rules to rule out the “de_2” by putting a value “2”to the attribute “level” of “的(de)” when the words before “的(de)” were adjective, quantifier, verb and pronoun;
- building 1 rule to identify the “de_1” by putting the “de_1” with a tag “MK”;

step2: recognizing the front and right boundaries

- recognizing the front boundaries L1 when the words were prepositions
- putting a tag “MK_Q” on the left boundary words when the words were nouns, NUM or SPN before the left boundary words;
- recognizing the rear boundaries L1H when the words were locatives etc;
- putting a tag “MK_H” on the right boundary words when the words behind the right boundary words were noun;

step3: Generating the PPs and NPs

- generating the PPs by combining the words by the combination pattern[3] of boundary: (L1,L1H) (L1,L1] ,

[3] Combination pattern (A,B) and [A,B] indicated combining the word from A to B, and “(“indicate A or B was included, and A or B was excluded.

[L1H,L1H), (L1,MK], [MK,L1], (NP_Bng%,L1H), [MK,L1H) by programming.

- generating the NPs from some combinations of [L1H,NP_End%), (MK,MK) by programming.
- generating the NP_Bs from some combinations of (NP_Bng%,L1] by programming and put an attribute "NOT_CHANGE".

Through three steps above, we could obtain all the translation units "PPs, NPs and NP_Bs" for reordering.

3.2 Reordering

In reordering phase, we defined the head NP at the end as NP_E. Then, the NPs contained PPs, NP_Bs, DE1, NP and NP_E. The strategies for reordering the Chinese NPs are as follows:

Step1: building 1 reorder rule for the Chinese NPs "NP_B PP NP";

Step2: keeping the NP_Bs not moving and move the PPs behind the NP_Es in reverse;

Step3: moving the other NPs behind the NP_Es in reverse, delete the "的(de)" and add the English word "of".

4 Experiments and Results

In order to test the result of this rule-based method and the strategy of reordering, the experiments takes 500 authentic patent texts provide by SIPO as the training set. The evaluation will use the development data for the NTCIR-9 Patent Machine Translation Pilot Task, containing 2,000 bilingual Chinese-English sentence pairs. After integrating the method into an existing rule-based system (HSCTMT), we take a closed test on training set and an open text on evaluation set.

Table 7. Accuracy of Translation Units

System	Precision (%)
Closed test	99.26
Open test	98.77

In table 7, the accuracy of translation units using boundary words in our system was very high and the results illustrated our rule-based method was efficient.

Table 8. Accuracy of Reordering of Chinese NPs in our system and Google

System	Closed test Precision (%)	Open test Precision (%)
HSCTMT	97.26	89.77
Google	57.47	59.92

In table 8, the result of two test shows the strategy of reordering was efficient, the semantic analysis in a rule-based method has effectively improved the recognition result of units for reordering, and Google performs poorly in tests. There are two factors affecting the performance:

- Incorrect boundaries of the nested PPs and PPs with a boundary affected the results.
- The multi-category words affected the results.

5 Related Works

Many reordering methods or strategy have been proposed in recent years to address this problem from different aspects. Phrase-based models excel at capturing local reordering phenomena and memorizing multi-word translation [2], but they perform poorly in the long and nested sentences in Patent. Syntax-based models handle long-distance reordering better than phase-based models. Reference [3] introduced a set Syntax-based rules to decide if a DE construction should be reordered or not before translating to English. Reference [4] focused on a Chinese noun phrase [A DE B] and explored a log-linear DE classifier by using syntactic, semantic and discourse context to producing an English translation strategy. Reference [1] explored a Hierarchical Semantic-Category-Tree (HSCT) model, which present a sentence as a hierarchical structure based on the Hierarchical Network of Concepts theory(HNC theory) and handle the reordering in three levels: Sentence Level, Chunk Level and Word Level. Reference [5,6,7] designed a Chinese-English Patent Machine Translation system based on the HSCT model.

6 Conclusions and Future Work

Based on analysis of translation units, we used a rule-based method to recognize the boundaries of the translation units using boundary words. Based on the analysis Chinese-English orders of Chinese NPs with two or three translation units, we developed a strategy on how to reorder the Chinese NPs. The experimental results showed that our rule-based method and strategy were very efficient on the reordering the NPs.

In future, we will enrich and refine the rules to improve the performance and research on how translate Chinese NPs fluently.

Acknowledgement

The authors are grateful to Multi-level Knowledge Representation of Mass Texts and Chinese Language Understanding System (National 863 Program, No.2012AA011104) and the Fundamental Research Funds of Central Universities for financial support.

Reference

[1]Zhu Xiaojian, Jin Yaohong,2010, Hierarchical Semantic-Category-Tree Model for Chinese-English Machine Translation, China Communications, Beijing,80-92

[2]KOEHN P,MARCU D. Statistical Phrase-based Translation [C]//Proceeding of the 2003 Conference of the North American Chapter of the A ssociation for Computational Linguistics on Human Language Technology, 48-54.

[3]Chao Wang, Michael Collins and Philipp Koehn. 2007 , Chinese syntactic reordering for statistical machine translation. In Proceeding of EMNLP-CoNLL, Prague, Czech Republic, June. Association for Computational Linguistics, 737-745.

[4]PiChuan Chang, Dan Jurafsky and Chrisstopher D. Manning.2009, Disambiguating "DE" for Chinese-English Machine Translation. In Proceeding fo the Fourth Workshop on Statistical Machine Translation, Athens, Greece, 30 March-31 March. Association for Computational Linguistics. 215-223.

[5]Yun Zhu, Yaohong Jin.: A Chinese-English Patent Machine Translation System Based on the Theory of Hierarchical Network of Concepts .J: The Journal of China Universities and Tele-communications, Beijing, 140－146. (2012)