An Experiment of Machine Translation of Chinese Double-object Construction into English based on Semantic-selection Translation Model

Xu Dekuan[1]and Zhang Guohua[2]

[1]School of Literature
Ludong University
Yantai City, Shandong Province, China, 264025
xudekuan@tom.com; dekuan68@sina.com

[2]School of Foreign Languages
Zhongnan University of Economics and Law
Wuhan City, Hubei Province, China, 430073
huntergh@126.com

ABSTRACT. *Eleven Chinese double-object sentences with different thematic combinations were translated automatically with "Yaxin CAT 2.5", a commercialized Chinese-English machine translation system. The results indicate that the system failed to deal with most of the sentences. Then taking advantage of the self-defining grammar function of this system, grammatical patterns of translating Chinese double object sentences into English were added to the system but the translated version were still not satisfactory. At last, with the "class" technique of the system, machine translation models based on semantic selection were added to the system and relatively satisfactory automatic machine translations were produced by the system.*

Keywords: Double-object Construction; Machine Translation Model; Semantic Selection; "Class" Technique

1. **Introduction.** Double-object construction (DOC) is one of the important syntactic constructions in Modern Chinese and translating sentences of this construction is one of the difficulties in machine translation systems. The authors first investigated the treatment of DOC in "Yaxin CAT 2.5", a commercialized machine translation system and found that the results were far from being satisfactory either using the grammatical rules of the system or without using any grammatical rules. "Yaxin CAT 2.5" offers a "class technique" in the system and the users can add machine translation rules to the system and can choose to use the user-defined translation rules or not in the process of translation. Machine translation rules using "class technique" can be written based on semantic selection properties of different verbs. An experiment of was made on translating Chinese double-object construction into English based on semantic-selection translation model with "Yaxin CAT

2.5". The experiment process and results are as follows.

2. Examples of and Analysis on Machine-translated Chinese Double Object Constructions into English

2.1. **Chinese double object sentences and their English translation.** Xu (2004) provides 11 types of Chinese double object constructions in accordance the thematic combination of the two objects. We first write out them in sentences and offers a human-translated English version for reference (table 1), then put them in a machine translation system for automatic translation and make an analysis and evaluation of the machine-translated results.

TABLE 1. Chinese double object sentences for machine translation test and human-translated English version for reference

No.	Thematic combination	Chinese sentence	Human-translated English version for reference
1	source ＋ theme	你买张三一本书	You buy a book from Zhang San.
2	goal＋ theme	你递张三一块砖	You pass a brick to Zhang San.
3	beneficiary＋theme	你奖张三一元钱	You reward Zhang San with one Yuan.
4	beneficiary＋content	你补偿张三三天时间	You give Zhang San three days for compensation.
5	malficiary＋patient	你打张三一个碗	You break a bowl of Zhang San.
6	patient＋theme	你偷张三一块手表	You steal a watch from Zhang San.
7	patient＋content	你浪费张三五天时间	You waste five day's time of Zhang San.
8	object＋result	你吐张三一口痰	You spit at Zhang San.
9	object＋theme	你介绍张三一个朋友	You introduce a friend to Zhang San.
10	object＋content	你报告张三一个好消息	You report a piece of good news to Zhang San.
11	object＋title	你叫张三约翰	You call Zhang San John.

2.2. **The machine translated results of "Yaxin CAT 2.5" and analysis.** "Yaxin CAT 2.5" is commercialized software developed by Yaxincheng Software Company of Beijing (Yaxincheng Software Company of Beijing 2000). Both the grammar library and lexical library of "Yaxin CAT 2.5" are open to the user: the user can made changes to the grammar library and lexical library; the user can choose to use or not to use the grammar library and lexical library provided by the "Yaxin CAT 2.5" in the process of machine translation. With these characteristics, we can make machine translation experiments in "Yaxin CAT 2.5".

First, let's look the automatic translation results without any grammatical rules (table 2).

TABLE 2: Machine translation results by "Yaxin CAT 2.5" without any grammatical rules

No.	Chinese sentences	Translation by Yaxin CAT 2.5
1	你买张三一本书	You bought Zhang San a copy of book.
2	你递张三一块砖	You past Zhang San a piece of Brick.
3	你奖张三一元钱	You award Zhang San one Yuan money.
4	你补偿张三三天时间	You Compensation Zhang San hour.
5	你打张三一个碗	You Dozen Zhang San one bowl.
6	你偷张三一块手表	You Stole Zhang San a piece of Wristwatch.
7	你浪费张三五天时间	You Waste Zhang San penthemeron hour.
8	你吐张三一口痰	You spat Zhang San mouthful sputa.
9	你介绍张三一个朋友	You introduce Zhang San one friend.
10	你报告张三一个好消息	You Report Zhang San one hot wire.
11	你叫张三约翰	You cry Zhang San John.

From table 2 we can see that, without grammatical patterns, the system simply translates Chinese into English word for word and chooses English word according to the first definition in the Chinese-English dictionary of the system, therefore many problems can be found in the translation. For example, in the sentence No. 5, the Chinese verb "打" is translated into an English noun "dozen". Such a translation is far from being satisfactory.

We know look at the translation with the grammatical patterns provided by the system (table 3).

TABLE 3: Machine translation results by "Yaxin CAT 2.5" without any grammatical rules

No.	Chinese sentences	Translation by Yaxin CAT 2.5
1	你买张三一本书。	You bought Zhang San a copy of book
2	你递张三一块砖。	You past Zhang San a piece of Brick.
3	你奖张三一元钱。	You award Zhang San one Yuan money.
4	你补偿张三三天时间。	You the recovery for Zhang San hour.
5	你打张三一个碗。	You Dozen Zhang San one bowl.
6	你偷张三一块手表。	You Stole Zhang San a piece of Wristwatch.
7	你浪费张三五天时间。	You dally Zhang San penthemeron hour away.
8	你吐张三一口痰。	You spat Zhang San mouthful sputa.
9	你介绍张三一个朋友。	You introduce Zhang San one friend.
10	你报告张三一个好消息。	You report on Zhang San one hot wire.
11	你叫张三约翰。	You ask Zhang San John to.

Compared with table 2, the translations are the same except No. 4, 7, 10 and 11. Even for these four sentences, their English version is still far away from the human translation. For example, No. 11, the English translation is not a complete sentence. After investigation, we find that, for "买、递、奖、打、偷、吐、介绍", there are no grammatical patterns and for "补偿、浪费、报告、叫", the grammatical patterns provided by the system are as table 4.

TABLE 4: the machine translation grammatical patterns for "补偿、浪费、报告、叫" in "Yaxin CAT 2.5"

Chinese model	English model
补偿 *	make #1 up
补偿 *	compensate #1 for
补偿 *	the recovery for #1
浪费 *	squander #1 away
浪费 *	molder #1 away
浪费 *	fribble #1 away
浪费 *	fool #1 away
浪费 *	dally #1 away
报告 *	report on #1
叫 *	ask #1 to

The symbols "*" and "#" are those in "Yaxin CAT 2.5" used for defining translation grammatical patterns. The common syntactic patterns or grammatical patterns are centered on some words or some type of words. In the actual usage of language, what to be changed are some words or word groups. "Yaxin CAT 2.5" takes advantage of this characteristics of language and makes some expansion in order to help in machine translation: for the fixed parts of the sentence, no changes are to be made both in the original text and translated text; for the variable parts, "*"is used in the original text and "#" for the corresponding translation. In the process of machine translation, the system will analyze the original sentence and compare it with the defined grammatical patterns, if there is/are some grammatical pattern(s) matched, the system will replace "*" in the original into "#" in the translated text. The number after "#" indicates the translated words corresponding to the number of "*" in the original text: "1" for the first"*", "2" for the second and so on. For example, the first rule for "补偿" →" make #1 up ", in translation, the system will put "补偿" into "make…up", and render what after "补偿" into English and put the English between "make" and "up". For example, "补偿损失"will be translated into "make the loss up". From table III we can see that for "补偿" there three rules, and for "浪费" five and for "报告" and "叫" one respectively. However, all these rules are for one element after the verb and are not good to deal with double object constructions.

In one word, for the 11 sentences, either there is no grammatical pattern at all or the

existing rules are for one element after the verb. So for double-object sentences, the system cannot translate them rightly. Therefore, proper grammatical patterns dealing with double object-sentences should be added to the system for translating them.

3. **Grammatical Models for Machine Translation of Chinese Double Object Construction into English**

3.1. **Grammatical models for machine translation of Chinese double object construction into English.** We write out the translation rules following "Yaxin CAT 2.5" convention (Table 5).

TABLE 5: Grammatical patterns for translation some Chinese double object sentences into English

No.	Chinese sentences	Chinese model	English model
1	你买张三一本书。	买**	buy # 2 from # 1
2	你递张三一块砖。	递**	pass #2 to # 1
3	你奖张三一元钱。	奖* *	reward #1 with #2
4	你补偿张三三天时间。	补偿* *	give # 1#2 for compensation
5	你打张三一个碗。	打* *	broke #2 of #1
6	你偷张三一块手表。	偷* *	steal #2 from #1
7	你浪费张三五天时间。	浪费**	waste #2 of #1
8	你吐张三一口痰。	吐**	spit at #1
9	你介绍张三一个朋友。	介绍**	introduce #2 to #1
10	你报告张三一个好消息。	报告**	report #2 to #1
11	你叫张三约翰。	叫**	call #1 #2

The 11 sentences in Table 5 represent the 11 types of thematic combinations of the two objects in double object construction. As we all know, the morphological changes in modern English is not evident so the thematic roles are not reflected in morphological endings: some thematic roles are reflected by different prepositions, whereas the most thematic roles have no special marks, just as the case in Modern Chinese.

There is another interesting phenomenon in English: verbs with similar meanings may use different prepositions, i.e., similar thematic roles are expressed by different prepositions. For example, when Chinese words "偷" and "抢" are used in double object construction VN_1N_2, the thematic roles for N_1 in "偷 N_1N_2" and "抢 N_1N_2" are both "malficiary" and the thematic roles for N_2 in the relevant constructions are both "theme". Whereas in English, one must say "steal something from somebody" and "rob somebody of something" respectively: the "theme" role for "steal" is expressed with no preposition and that for "rob" is expressed by the preposition "of", while the malficiary role is expressed by the preposition "from" for "steal" (the thematic role here is better understood as "source") and that for "rob" is expressed with no preposition. Therefore, the English pattern in the table is not given in accordance with thematic roles in Chinese, but according to the concrete

circumstances.

3.2. **The machine-translated results after the grammatical models are added to the system**

We add the rules in table 5 to the system and get the translated results as in table 6.

TABLE 6: The results after using the grammatical patterns

No.	Chinese sentences	Translation by *"Yaxin CAT 2.5"*
1	你买张三一本书。	You buy from Zhang San a copy of book.
2	你递张三一块砖。	You pass to Zhang San a piece of brick.
3	你奖张三一元钱。	You reward with Zhang San with one Yuan Money.
4	你补偿张三三天时间。	You give for compensation Zhang San three day time.
5	你打张三一个碗。	You break of Zhang San one bowl.
6	你偷张三一块手表。	You steal from Zhang San a piece of wristwatch.
7	你浪费张三五天时间。	You waste of Zhang San five day time.
8	你吐张三一口痰。	You spit at Zhang San.
9	你介绍张三一个朋友。	You introduce to Zhang San one friend.
10	你报告张三一个好消息。	You report to Zhang San one good news.
11	你叫张三约翰。	You call Zhang San John.

After comparing the machine translated version with the reference translation, we find that, only the 8th and 11th sentences are desirable and the rest 9 sentences are problematic. For example, the first sentence, "你买张三一本书", the grammatical pattern we offer is "买**"..."buy # 2 from # 1". There two words after "买": "张三" and "一本书", according to the translation pattern, the second word "张三" should be translated into the corresponding English word "Zhang San" and come after "from" and the second word "一本书" should be translated into the corresponding English word "a book" and come after "buy", thus resulting in a good translation such as "You buy a book from Zhang San". But the machine translated result is "You buy from Zhang San a book". After analysis, we guess that it is difficult to deal with two succeeding "*"s in the grammatical pattern: it only recognizes one"*" and ignores all the other "*"s after the first one, therefore "#2" corresponding the second "*" is not effective.

4. The machine translation models based on semantic selection

4.1. "Class technique" and translation models for double object sentence based on semantic selection. In addition to grammatical patterns, "Yaxin CAT 2.5" provides "class technique". "Class" refers to word class or semantic attributes. Through "class technique" the system can deal with a set of sentences with common word class or semantic attributes involved. As far as double-object construction is concerned, the indirect object has a common property in semantic selection: a personal noun or personal pronoun should be used as indirect object. Thus we can take advantage of "class technique" to define translation models for double-object sentences. The translation models we offer are in table 7.

TABLE 7: Translation models for double object sentence based on semantic selection

No.	Chinese sentences	thematic combination	Chinese model	English model
1	你买张三一本书	来源＋客体	买{prn} *	buy #1 from @1
2	你递张三一块砖	目标＋客体	递{prn} *	pass #1 to @1
3	你奖张三一元钱	受益者＋客体	奖{prn} *	reward @1 with #1
4	你补偿张三三天时间	受益者＋内容	补偿{prn} *	give @1 #1 for compensation
5	你打张三一个碗	受损者＋受事	打{prn} *	break #1 of @1
6	你偷张三一块手表	受损者＋客体	偷{prn} *	steal #1 from @1
7	你浪费张三五天时间	受损者＋内容	浪费{prn} *	waste #1 of @1
8	你吐张三一口痰	对象＋结果	吐{prn} *	spit at @1
9	你介绍张三一个朋友	对象＋客体	介绍{prn} *	introduce #1 to @1
10	你报告张三一个好消息	对象＋内容	报告{prn} *	report #1 to @1
11	你叫张三约翰	对象＋称呼	叫 {prn} *	call @1 #1

In the table, for the Chinese model, what is in "{ }" indicates the semantic selection of the indirect object, here "prn" standing for "person noun", the number after "@" in the English model corresponding the relevant element in "{}". The significance for "*" and the number after "#" is the same as above. Take the first one as an example: "买 {prn} *" →"buy #1 from @1". This means that the English model corresponding to Chinese "买＋person noun＋ other word(s)" is "buy … from …": when the system meets "买＋ person noun＋ other word(s)", in accordance with this rule, the system will translate "买" into "buy", translate other word(s) after the person noun into English and put it immediately after "buy" and translate the person noun into English and put it after "from".

4.2. The machine translation results after the models based on semantic selection applied in the system.
We add the rules in table 7 to the system and get the translated results as in table 8.

TABLE 8: The results after using the translation models based on semantic selection

No.	Chinese sentences	Translation by *"Yaxin CAT 2.5"*
1	你买张三一本书。	You buy a copy of book from Zhang San.
2	你递张三一块砖。	You pass a piece of brick to Zhang San.
3	你奖张三一元钱。	You reward Zhang San with one Yuan Money.
4	你补偿张三三天时间。	You give Zhang San three day time for compensation.
5	你打张三一个碗。	You break one bowl of Zhang San.
6	你偷张三一块手表。	You steal a piece of wristwatch from Zhang San.
7	你浪费张三五天时间。	You waste five day time of Zhang San.
8	你吐张三一口痰。	You spit at Zhang San.
9	你介绍张三一个朋友。	You introduce one friend to Zhang San.
10	你报告张三一个好消息。	You report one good news to Zhang San.
11	你叫张三约翰。	You call Zhang San John.

5. **Discussion of the results.** After comparing the automatic machine-translated version with the human-translated version, we find that the machine translated version is quite pleasing except for some small draw-backs: for example, in No. 7 "five day time" and in No. 10 "one good news" are not quite good English phrases. This is due to the fact that when translating Chinese "Numerical + Classifier + Noun" construction into English, complicated and detailed rules should be added to the machine translation system. As far as double-object construction is concerned, the machine translation is satisfactory.

6. **Conclusion and Future works.** Double object construction is one of the difficulties in Chinese-English machine translation. We made an experiment with 11 Chinese double object sentences in "Yaxin CAT 2.5" machine translation system and after machine translation models based on semantic selection are added to the system relatively satisfactory automatic machine translations are obtained.

We used only 11 sentences with different thematic combinations in this experiment. As we noted in section 3.1, as for the English translation, the uses of prepositions are more varied. Therefore, more types of sentences are to be investigated in the future.

REFERENCES

[1]   Xu, Dekuan. A Study on Chinese Double-object Construction [M]. Shanghai: Shanghai Dictionary Press. 2004.

[2]   Yaxincheng Software Company of Beijing. Yaxin CAT User Manuel [M]. Beijing: Beijing Hope Electronics Press. 2000.

[3]   Yang, Muyun et al. Machine Translation System [M]. Harbin: Harbin Institute of Technology Press.2000.

[4]   Zhao, Tiejun et al. Machine Translation Principles [M]. Harbin: Harbin Institute of Technology Press. 2000.

[5]   Chao, Yuen-ren. A grammar of Spoken Chinese [M]. University of California Press. Berkeley. 1968.

[6]   Larson, R. "On the double object construction" [J]. Linguistic Inquiry 19. 1988.

[7]   Chomsky, N. Aspects of the Theory of Syntax [M]. Cambridge: Cambridge University Press. 1965.