

A Target-oriented Case Study on Some Chinese Syntax-based Semantic Restrictions

Xiaohong Wu

Centre Tesniere, Faculty of Letters,
University of Franche-Comte, France
Faculty of Foreign Languages,
Minzu University of Qinghai Province,
China
wuxiaohongfr@yahoo.com.cn

Sylviane Cardey

Centre Tesniere, Faculty of Letters,
University of Franche-Comte, France
sylviane.cardey@univ-fcomte.fr

Abstract

Concerning the relationships between a verb and its complements much work has been done in recent years, especially those between a verb and its nominal and prepositional complements. However, another fundamental yet not deeply studied problem for the Chinese language in the application of machine translation (MT) is that the relationship between a verb and its complements of other grammatical categories, e.g. adjectives, still remains a big challenge. Such relationships are often related to certain particular syntactic structures and are thus bound by syntax-based semantic restrictions. In the case of Chinese linguistics, the *ba*-construction and its counterpart structure manifest exactly such a kind of phenomenon. In this paper we demonstrate the problems we encounter by illustrating the syntactic and semantic variations produced by these structures. We explain in detail how, if there is a lack of such syntactic and semantic information, the MT quality is influenced by generating badly constructed Chinese sentences which are neither acceptable nor grammatical. Since we apply controlled language (CL) technique as an aid to our work, we narrow our research to limited syntax and restricted semantics.

1 Introduction

In this paper we describe the work we do for the LiSe project¹, focusing on machine translation in the domains where safety is extremely crucial and the translation of such texts thus requires a high degree of accuracy. Due to the characteristics of the project, we start the research work first from the design and development of a controlled language (CL) in the hope that taking the advantages of CL we could greatly diminish the syntactic and semantic complexity and ambiguity from all levels for the languages with which we work with. We select three sub-domains for the CL applications: medical protocols, public warnings and certain emergency messages (for more information, see Wu, et al. 2006; Wu, 2005). Texts of this kind in our corpus are similar to those of users' manuals or practical instructions and thus are relatively easier to reduce in size or make structurally simpler and semantically less ambiguous. All the texts are translated from French into four target languages: Arabic, Chinese, English and Thai by human translators and verified by domain specialists. By doing so, we successfully build several parallel corpora for our language comparison and linguistic analysis, taking French as the source language (SL) and the others as the target languages (TL).

These parallel corpora are then distributed to different linguists of each language for re-editing, that is, to apply CL technique to make the sentences in the text shorter and simpler using the same rewriting guidelines. The re-edition of the texts is often done with the intervention of the domain specialists so as to make the sentences not only structurally simpler but also easier to understand. Except for English and

¹ LiSe (Linguistique et Sécurité), project funded by the ANR (French National Research Agency) contract: ANR-06-SECU-007.

French which are relatively more close to each other, the other three languages are linguistically quite far from these two and from each other. Without the help of CL techniques it would be undoubtedly a big challenge. In this paper, we present as an example only part of the work performed on three of the languages so as to demonstrate the problems produced by the linguistic discrepancies among them: Chinese, English and French. In addition, our presentation also centers on two basic Chinese structures – the ba-construction and the cohesive complement to show how problems arise if there is a change of the structure or the word order.

2 Language particularities and problems faced

Discussions on ba-construction have remained one of the most important topics in Chinese linguistics circles. The enormous number of results provided by an Internet search on the topic might give us an impression that there is no room for anyone to continue on this topic. It is true that in relation to the possible constructions of the structure, no matter from a syntactic or a semantic point of view, almost all facets of this construction have been explored both diachronically and synchronically. However, much of such linguistic research is found to be difficult to be integrated directly in the application of MT. Consequently, this frequently used Chinese syntactic construction is seen to be very badly translated into or, especially from other language(s). An obvious evidence for this bad performance is that the ba-construction is by itself a Chinese language specific phenomenon which finds no place in other languages. Besides, this superficially simple structure (ba + NP + VP) in fact covers almost all the types of Chinese phrasal constructions and also owns some semantically very complicated particularities according to the varied ways it is employed, of which the latter is still far beyond the description of present computational linguistic knowledge. Furthermore, the semantic content of this construction may vary if any change of the word order in the syntax happens. Therefore, the difficulty lies directly in the generation of the Chinese equivalents from other language(s) if the ba-construction is mandatory in interpreting the meanings conveyed in the source language (SL), such as in our case.

In our work we firstly build our own specific lexical information and grammar rules for the creation of this construction while transferring a French sentence into a Chinese one which needs this construction. Our analysis is based on the systemic or micro-systematic linguistic analysis theory which stresses the fact that a model ought to serve not just for the description of a precise phenomenon but also as a resource allowing solutions to other linguistic phenomena covered by the model as well as the linguistic phenomena not yet identified for which the model and/or others could use. The theory advocates the decomposition or recomposition of linguistic phenomena in order to analyze them better. For example, instead of listing all the elements which make up part of a language, it is better to try to classify, order, arrange or group them so as to define if certain of them can function as a complete independent system or as interrelated systems according to what ought to be demonstrated or solved. The theory is by itself based on set theory. When confronted with a problem or a precise phenomenon we try to choose the necessary elements and structure them in the form of a system. In this way, the problem can be manipulated as it is apprehensible. Generally, lexis, morphology, syntax and semantics cannot be separated when one deals with language phenomena. It is thus necessary to: identify the problem to be solved; construct the system or the systems which can be interrelated; describe then the problem using the system/s in order to solve the problem where necessary. Applications of the theory can involve different languages, different problems and different domains (Cardey, 2009).

While defining the syntactic rules both for the phrasal level and for the sentence level, we came across some difficulties caused by the language particularities even in some basic and simple structures. For example, imperative sentences take an important role in our corpus and this kind of sentences has to be translated into Chinese in most cases by generating a sentence containing the ba-construction. Therefore, we need to mark this information directly on the verbs which can be used in this structure, such as [$\pm$ ba]. However, some verbs can be used in both ways, that is, either with the ba-construction or without it. It seems easy to add a rule indicating both options. The problem here is that this seemingly easy technique

is often jeopardized by some syntactic-semantic relationships inhabited within this structure, which is quite different from that of the English and French counterparts.

2.1 Language particularity – the ba-construction

Superficially, the ba-construction has a simple structure which can be represented as ba + NP + VP. Zhan (2004) classifies the types of both NP and VP compositions in the ba-construction from two aspects, one from the phrasal syntactic types of VP and the other from the semantic role types that the NP can take. He therefore classifies the syntactic construction of VP into 12 types and the semantic roles of the NP into 9 types. In his paper Zhan also offers the statistical figures calculated in relation to the ba-construction. He took the Modern Chinese Grammar Information Dictionary (published in 1997) as a resource for this work. According to his investigation, this dictionary lists 10,283 verbs among which 5,933 can take nominal complements. 1,803 verbs out of all these verbs can be used in ba-construction, which according to his further examination should be more than that as he found out that some verbs that otherwise can be used in this construction are not so marked. However, all these figures do not provide us with any immediate inspiration in that we have a limited list of verbs and syntax to worry about. Unlike European languages, the Chinese language does not possess a large vocabulary but it has a strong capacity for word formation and many words are occasionally formed or can be framed for a temporary use with added unpredictable semantic information. This might produce some degrees of influence for both the syntax and the semantics of a given sentence. This information is hard to catch and much of it is inevitably far beyond our present scope of research.

Generally speaking, only transitive verbs can be used in this construction. Therefore many sentences with a transitive verb can be transferred into a sentence with the ba-construction. That is, sentences that are of SVO sequence can be changed into ones with the ba-construction. When a sentence which has the surface structure of SVO is transferred into the ba-construction, some changes take place to the general structure. The object (usually a NP) of the verb is shifted to the front of the verb which makes the whole SVO structure possesses the surface structure of an SOV sequence, having the NP become the complement of the preposition ba.

Another fact is that ba as a preposition lacks semantic content by itself. However, it constrains the selection of verbs as we have mentioned. In other words, not all transitive verbs can be used in this construction. According to Ye, et al. (2007), the whole construction has “abstract meanings” as “disposal” or “causation” which are independent of the content words inhabiting it. “Therefore only transitive verbs that encode such meanings are permitted to appear in this construction”. They further stress that ‘a verb which is perfectly acceptable in the SVO structure but does not have the appropriate “disposal” or “causation” meaning would constitute a construction-based semantic violation if it is forced to appear in the ba-construction’. They refer to this as a construction-based semantic constraint which cannot be attributed to the word ba.

Another equally important phenomenon closely related to this structure is that after the object of the transitive verb has been shifted to the front and a ba-construction is formed, many verbs by themselves become semantically incomplete and thus require another complement to fill the slot left behind by the shifted object. This newly obtained complement is often made up of words of other grammatical categories or a phrase. This new complement often adds some more semantic or even pragmatic nuances to the verb and/or the sentence to indicate different levels of information associated with the action. Such nuances differ from those expressed in their SVO counterpart.

Interestingly, many such syntax-based semantic violations can be resolved by adding a post-posed complement(s) to the verb and result in the sentence being acceptable and grammatical. That is to say, if a verb cannot satisfy such a semantic requirement, it cannot be used in this construction, or we have to add more constituent(s) to this construction to satisfy this syntax-based semantic restriction.

Taking the examples of Ye et al. as ours, we can see that although the verb “欣赏 (close to ‘enjoy/appreciate or view’)” cannot be used in the ba-construction as in the example below:

- a) *市民把名画欣赏 (LWS²: citizen ba famous picture enjoy/view).

we can make it canonical by adding a post-posed complement such as a verbal classifier. To make the sentence read more naturally we change the subject to a pronoun (我, I).

a.1 我把名画欣赏了一遍 (LWS: I ba famous picture enjoy/view once³; I enjoyed/viewed the famous pictures thoroughly).

By doing so, we have well satisfied the semantic ‘need’ for “causation”.

In Chinese linguistics, this kind of postposed complement is often divided into two different kinds according to its syntactic behavior, of which one is called the cohesive complement and the other is called the compositive complement. Here lies our major problem: the interpretation and generation of the cohesive complement of the verb.

2.2 Language particularity – adverbial vs. cohesive complement – the problems faced

In all languages, for a certain meaning there often exist varied ways to express this, that is, one meaning can be expressed differently using different words and/or different syntactic structures (Milićević, 2006). Generally speaking, if a meaning is articulated using different ways, some semantic nuances might thus be produced. It is also possible that the general meaning is the same in the differently expressed sentences but that there might be some kind of syntactic restrictions for such different usages.

For example, in all these three languages: French, English and Chinese, some of the verb adjuncts (the constituents that modifies the verb) can appear both in front of the verb and after it. In Chinese, however, if a verb takes an adjunct that appears in front of the verb, we generally refer to it as the adverbial of the verb, which is in fact the most common position of the adverbial in Chinese. If the adjunct appears after the verb, it is considered as another kind of syntactic category hence forward called the cohesive/compositive complement. This kind of verbal complement is different from the adverbial and such differences in the syntax leads to the dissimilarity of the grammatical function. Like the ba-construction this kind of constituent belongs to the category of language specific phenomena. Therefore, while translating a sentence into Chinese, we often have to “replenish” this kind of constituent to the translated sentence so as to make up for the semantic incompleteness required by the syntax.

Sometimes the same Chinese word (usually an adjective or an adverb) can appear in both positions (not at the same time). This results in two different syntactic constituents and often produces some change in the semantic emphasis. The option of the position in the phrasal structure depends not only on the sub-categorization scheme of the verb but also on the kind of syntactic structure where it is allowed to appear.

We look at the following two groups of VP, in which the same word appears in two different positions and demonstrates exactly such a situation despite the fact that we have just the same word but in a reversed order.

The change of the word order in the following examples does not produce much semantic difference but will be syntactically restricted if further used in the context of a sentence.

Group 1	躺 _v 平 _{adj}	坐 _v 端 _{adj}	放 _v 平 _{adj}
Group 2	平躺	端坐	平放
Translation	s'allonger sur le dos ?? lie on one's back	s'asseoir ?? sit still	poser à plat ?? lay flat

Table 1 Structure variations caused by word order

As we have just mentioned, the reversed word order results in two different syntactic structures: while the first one is a verb plus an adjective forming a cohesive complement; the second can be regarded as an

² LWS: short for: literal word sequence

³ “Once” here could also be interpreted as thoroughly or one after another

adverbial. Both of them can function as the predicate in a sentence. However, while the first group can be used as an independent sentence such as an imperative sentence, the second one when used in the similar way has to be supplemented with another constituent – a cohesive complement to indicate accordingly certain semantic needs as seen in the following example.

b) 请坐端 (Please sit still)。

However with the second choice, it has to be:

c) 请端坐在椅子上(LWS: please still sit on chair; Please sit still on the chair)。

This is instead of (d) which is ungrammatical:

d) *请端坐 (LWS: please still sit)。

This means that a verb having the cohesive complement tends to be semantically self-sufficient while a verb with the preposed adverbial in the same context is not. In other words, if we use the second group of words as the predicate in a sentence with the ba-construction, without the interference of a cohesive complement, it violates the syntax-based semantic restriction. This reflects the fact that the semantic restriction is not only required by the lexical unit but also by the syntactic structure. This also explains the reason why in some cases, this kind of violation might occur and/or such violations can be resolved by adding a cohesive complement. In Example (c), the prepositional phrase “on the chair” assures the semantic content of “location”. As a result, there is no violation for (c) but there is so for (d).

It is clear that for a VP with only an adverbial, this tends to be more easily restricted by the syntactic-semantic relationships. However, simply reversing the position of the adverbial with the cohesive complement cannot always be the right solution as the actual situation is more complicated than that. This complication is often increased by the selection of the word(s) that can take the position of the adverbial or of the cohesive complement. To be brief, a word that can take the position of an adverbial does not necessarily mean that it can also take the position of a cohesive complement, and/or vice versa, though many of them might seemingly present almost the same syntactic or semantic resemblance. For certain of such phrases if we reverse the order, we will either create a structure which does not make any sense or a structure with quite different semantic content. See the examples in the following table.

Group 3	用(to use)坏(bad) : to wear out	倒(to pour)完(over): to pour out	变(to turn)冷(cold): to turn cold
	*坏用	*完倒	*冷变
Group 4	擦 _v 干 _{adj} (to dry)	学 _v 好 _{adj} (to learn well)	弄 _v 白 _{adj} (to rub up or to whiten sth)
	干擦 (to rub/erase/clean with sth dry; or to rub in vain)	好学 (easy to learn)	白弄 (to do sth in vain)

Table 2 Examples of reversed word sequence

Words in Group 3 are exactly like those in Group 1: a verb plus an adjective. However, they cannot be used in a reversed way as shown in the row that follows. These reversed counterparts are not even the recognized Chinese words. The words in Group 4 when used in a reversed way do not have the same meaning at all, as the interpretations indicate.

Here the problem is manifest: if a system cannot distinguish an adverbial from a cohesive complement, or it lacks the capacity to deal with such phenomena, it will fail to generate the right sentences as is often seen in MT. This fact has caused us to start the study for the cohesive/compositive complement which until now appears to have been ignored.

Let us see a real example from our own MT system.

French	English
e) Laver bien toutes les bouteilles.	Wash well all the bottles.
Chinese	
e.1 *把所有瓶子 <u>好</u> 洗。(LWS: ba all bottle well wash)	

Using our own MT system, we first get a bad Chinese translation (see e.1) though it correctly matches the suitable lexical and syntactic equivalents that are predefined in the lexicon and in the grammar database. The MT system well generates the ba-construction as we have very efficient linguistic rules to deal with this structure. However, it makes a wrong judgment by taking “好” as the adverbial of the verb. In fact in this context “好” cannot take the position of the adverbial. If this word happens to appear in this position, it should be used in its duplication form, i.e. AA + V form, such as “好好 + V”. For example, “把字好好写 (LWS: ba word well well write; Write carefully the words.)”, “把话好好说 (LWS: ba speech/talk well well speak; Speak slowly/clearly/patiently <what you want to say>.)”.

Yet, if “好” functions as the cohesive complement, it is well received. That is to say, when used in a “V + 好” structure, e.g. “把字写好 (LWS: ba word write well; Mind your handwritings <try to write beautifully>.)”; and “把话说好 (LWS: ba speech/talk speak well; Speak properly <what you want to say>.)”. The problem here is that the change of the position produces the change of the meaning too, as we can see in the interpretations.

In addition, both of the just mentioned syntactic options are neither grammatically correct nor semantically exact in our case example as follows:

e.2 *把所有瓶子好好洗。(LWS: ba all bottle well well wash) → syntax-based semantic violation (it can be released by further adding a verbal classifier “一下” after the verb “wash”).

e.3 把所有瓶子洗好。(LWS: ba all bottle wash well) → semantically inaccurate

(e.3) illustrates our concern. If we move the word “好” after the verb and make it become the cohesive complement, it is grammatically correct. Semantically the sentence transfers an idea that is a bit different from the SL. It means “wash the bottles and make them ready for use” while the original meaning should be “wash thoroughly all the bottles to make them clean”. This means that if we simply shift the preposed adjunct to a postposed cohesive complement, it does not always resolve the semantic problem. The best choice is either to choose another word which is both semantically and grammatically correct or to employ another kind of syntactic structure rather than the ba-construction. For example, we can choose a more exact word to substitute “好”: the word “干净” (clean). For example:

e.4 把所有瓶子洗干净。→ both grammatically and semantically correct
(LWS: ba all bottle thoroughly wash clean)

Or we use other syntactic structures:

e.5 洗干净所有瓶子。(LWS: wash clean all the bottles)

It seems that by selecting the right equivalent and authorizing both syntactic options we could easily resolve this problem. However, if we do so, other obvious problems might reappear. First the French word “bien” is by itself polysemous which contradicts our CL rule of “one word one meaning”. Second, it is nevertheless hard to control the use of this highly frequently used word in the SL and we cannot assign many possible equivalents to it. Besides, if we impose on the users by avoiding using this word in many cases, it will make the users very frustrated in looking for alternatives that are not always at hand and thus make our system unfriendly. Furthermore, even if “bien” could be well replaced by “well” in translating it

into English, it will never be the case in Chinese as it is almost impossible to predefine one or even two substitutes that are the equivalents of the SL verbs which happen to be modified by “bien” in French and “好” or “干净” in Chinese. Meanwhile, this approved substitute(s) could also satisfy both the syntactic and semantic restrictions as illustrated above.

In the case of (e.3) it can be regarded as an acceptable translation even if the meaning is not so accurate. For many other cases, it will not be so easily accepted semantically. Besides, some Chinese compound verbs by themselves convey such a meaning and do not need to find an equivalent. We look at the same example by changing the verb and by adding an NP modifier which is also one of our approved syntactic structures:

French	English
f) Bien sécher toutes les bouteilles sur la table.	Well dry all the bottles on the table.

Chinese

f.1 把桌子上所有的瓶子晾干。(LWS: ba table on all de⁴ bottle dry)

f.2 晾干桌子上所有的瓶子。(LWS: dry table on all de bottle)

Though the word “干” can also be regarded as an adjective, however “晾干” is considered as one word instead of VP. In (f.1) and (f.2), it is not necessary to translate “bien” into any Chinese word though there is no problem in English. If we add “好” either as the adverbial or as the cohesive complement, it will result in a sentence that is only semantically different:

f.3 把桌子上所有的瓶子晾干好。(LWS: ba table on all de bottle dry well; It is good to dry all the bottles on the table.)

f.4 好晾干桌子上所有的瓶子。(LWS: well dry table on all de bottle; <by taking certain measures> it becomes easier/more convenient to dry all the bottles on the table.)

The same situation might also come from the SL. That is to say, a SL word is semantically self-sufficient; however, while translating it into Chinese it is mandatory to have a cohesive/compositive complement which is required by the syntax. For example,

French
g) Premièrement, laver toutes les bouteilles. Sécher les bouteilles et ...
English
First wash all the bottles. Dry the bottles and ...
Chinese
首先, 把所有的瓶子洗 <u>一遍</u> 。把瓶子晾干并.....
(LWS: first, ba all de bottle wash one time. ba bottle dry and ...)

In the above example, the appearance of the cohesive complement (underlined) is mandatory. Therefore, we must add such information restricted by the syntactic-based semantics (which does not exist in the SL) to our system in order that when such sentences are processed, they can be transferred into grammatical ones. To solve these problems, we wish to stress three aspects: 1) extend the lexicon to include more semantic and syntactic information for both verbs and adjectives/adverbs; 2) define syntactic mapping rules to cover all the information needed; 3) establish a syntactic-semantic network to link all this information.

⁴ De: a structural particle showing the possessive relationship of the noun, similar to English “of”.

3 Building linguistic models

Up to now we have just given a very narrow and brief description of the basic problems connected to the ba-construction, the adverbial and the cohesive complement. In real texts these syntactic structures exhibit much more complicated variations of which many are not reported in this paper.

As we have already constructed our models for the generation of the ba-construction, we focus on defining other sets of special rules for the construction of the cohesive/compositive complement.

To do this work, we automatically extract the sentences containing the ba-construction and all the verbs that are surrounded either by an adjective or an adverb. We then concentrate on about 30 verbs which can be found in our corpus and which might produce such problems if badly interpreted.

3.1 Lexicalization

In translation, it is well known that the translation between languages often has to deal with the fact of one-to-many and/or many-to-one situations, which is not only the choice of words but also the choice of the syntax. For the human being this is an exciting challenge but also a piece of ‘recreation’ work. However, this is not the case for MT. Therefore, while searching for possible solutions, we often have to choose a roundabout way in order to minimize the difficulty. One way to solve this problem is to take the task of lexicalization so as to narrow down the semantic gaps between words/phrases from different languages.

In this work, in order to tackle the problems of the semantic incompleteness produced by the syntactic structure, one of our important methods is to have some relatively fixed structures lexicalized. For example, if a SL verb or a SL phrase has to be replaced by a Chinese verb(s) which has to be accompanied by a cohesive/compositive complement, sometimes also by an adverbial, we fix it as one lexical unit to assure the syntax-based semantic completeness. We list these as individual lexical units accordingly, ignoring the fact that in the TL they might be phrases instead of words. It should be noted that in our work the meaning of each word is restricted to no more than one. Unless it is extremely necessary, we do not define more meanings. In the case of necessity, we approve some words to have two or three meanings with strict semantic constraints but it is never allowed to exceed this limit.

Here are a few examples of the lexicalized Chinese equivalents.

French	English	Chinese
mettre en ordre	tidy up	收拾好
mettre ... à l'envers	put ... upside down	倒置/放
verser entièrement	pour out	倒空
s'allonger (sur le dos)	lie down/lie on one's back	躺平/平躺

Table 3 Examples of lexicalized Chinese equivalents

3.2 Syntax mapping rules

In order to have a set of the syntactic rules that include this information in the TL, the syntactic structures are classified into different categories which might be further divided into several sub-categories accordingly so as to include the possible reconstruction information of the approved syntax. This approach is again supported by set theory. Here are some simple examples of part of the approved syntactic structures. To make it easier to read, we only choose a few syntactically very simple examples to show the basic construction of some types of the approved phrasal structures.

Here are some of our approved syntactic structures in relation to the Ba-construction.

NP1 + ba +NP2 + VP, in which the VP could be:

1) $VP \rightarrow (PP) + V + PP$, e.g.

French: Placer le flacon sur la table. English: Put the flask on the table.

Chinese: 把烧瓶放到桌子上。(LWS: ba flask put to table shang- Loc⁵)

2) VP → V + CC⁶, e.g.

Fr: Laver bien les bouteilles. En: Wash well the bottles.

Ch: 把瓶子洗干净。(LWS: ba bottle wash clean)

3) VP → (AdP) + V + CC, e.g.

Fr: Vider complètement l'eau du seau. En: Pour out the water from the bucket.

Ch: 把桶里的水(彻底)倒尽。(LWS: <completely> ba bucket Loc <inside> de water pour <finish>)

4) VP → V + NP + CC (consisted by verbal classifier), e.g.

Fr: Répéter la première procédure. En: Repeat the first procedure.

Ch: 把第一个步骤重复一遍。(LWS: ba first procedure repeat one bian-classifier)

For the verbs that could not be used in the ba-construction, we also define the approved structures. Here we take only one of them as an example.

(NP) + VP, in which VP could be:

5) VP → AdP + V + NP + (PP), e.g.

Fr: Observer attentivement le changement de couleur de l'eau.

En: Observe carefully the change in the color of the water.

Ch: 仔细观察水的颜色变化。(LWS: carefully observe water de color change)

Of course there are still many other types of structures that could function either as an adverbial or as a cohesive/compositive complement as for instance, the verbal complement which consists of particles (“了”, “着”) and/or set expressions.

3.3 Semantic tags and equivalent candidates in the lexicon

In order that the system can make some wise and exact decisions while translating French sentences into Chinese ones, we also add more information to the lexical units in the lexicon. For example, we have already defined the syntactic information related to verbs to indicate what kind of syntactic structure a verb can have. We then list one or two candidate words that can function either as the adverbial or as the cohesive complement. Here are two simple examples showing how these pieces of information are linked through different linguistic databases.

Lexical entry	SynInfo	SemInfo	AdjInfo
Fr: observer Ch: 观察	±ba	<human behavior>	Attentivement; bien 认真+AdP/-CC ; 仔细+AdP/+CC
Fr: poser Ch: 放	+ba	<human behavior>	Doucement 轻+AdP/ -CC

Table 4 Examples of the lexical entry

In this table we only list the information to show whether a Chinese verb can be used in the ba-construction or not. “±ba” means that the verb can be used both in the ba-construction and the structure without it. “+ba” means that the verb can only be used in the ba-construction. Therefore if a verb is

⁵ Loc: a noun of locality signifying the location, often used together with a preposition, here “shang” means “on the surface of”.

⁶ CC: short for the cohesive complement

marked as “-ba”, it means it is not allowed to appear in the ba-construction. Similarly, while “仔细+AdP/+CC” means that this adjective can take both the adverbial position and the cohesive complement position, “认真+AdP/-CC” means that “认真” can only take the adverbial position in our presently approved syntax. This information together with information stored in the grammar rule base assures that our system makes the right choices while generating Chinese sentences.

4 Evaluation

The evaluation of the work is done by using our own system. The first reason for such a choice is that the ba-construction is often badly translated using other commercialized or online free MT systems. In fact in most cases, these systems seldom succeed in generating a grammatical Chinese sentence containing the ba-construction. As a result, we cannot collect useful information for the improvement of our own system. Secondly, even if some sentences are well translated, it is still hard to set any standard for the evaluation purpose. Besides, though the linguistic phenomena we present in this paper are very important for a better translation, these are still not well implemented by most MT systems. Therefore, our only choice is to adapt our own system to take over the task.

We set three standards for the evaluation: good, acceptable and bad for structure, word choice and meaning transfer. We select only twenty sentences for the evaluation, of which twelve have to be generated into Chinese by employing the ba-construction while the other eight sentences must be generated into sentences that could not use the ba-construction. Our result is nevertheless satisfactory with 98% accuracy compared with the first translation without these new pieces of linguistic information. In the first translation, almost all the sentences that need a cohesive complement are not correctly translated with an accuracy of less than 60%. However, we must stress here is that within a domain-oriented task with well controlled syntax and lexical meanings, such an accuracy is not surprising.

5 Conclusion

In this paper, we have explained part of the work that we have done for the MT of well controlled sentences from French to Chinese by illustrating the problems produced by certain language specific particularities. Based on our analysis of these linguistic phenomena, we have managed to build linguistic models to solve the problems in order to improve our MT quality.

References

- Cardey, Sylviane. 2009. *Controlled Languages for More Reliable Human Communication in Safety Critical Domains*. Proceedings of the 11th International Symposium on Social Communication, Santiago de Cuba, Cuba, January 19-23, 2009, pp. 330-335.
- Cardey, Sylviane. Greenfield, Peter. Wu, Xiaohong. 2004. *Designing a Controlled Language for the Machine Translation of Medical Protocols: the Case of English to Chinese*. Proceedings of the AMTA 2004, LNAI/3265, Springer-Verlag, pp. 37-47.
- Milićević, Jasmina. 2006. *A Short Guide to the Meaning-Text Linguistic Theory*. Journal of Koralex, Vol. 8, pp. 187-233.
- Wu, Xiaohong. 2005. *Controlled Language – A Useful Technique to Facilitate Machine Translation of Technical Documents*, Lingvisticae Investigationes 28:1, 2005. John Benjamins Publishing Company. pp. 123-131.
- Wu, Xiaohong. Cardey, Sylviane. Greenfield, Peter. 2006. *Some Problems of Prepositional Phrases in Machine Translation*. Proceedings of the 5th International Conference on Natural Language Processing, FinTAL, Turku, Finland, 2006, Springer-Verlag – LNAI 4139, ISBN 3-540-37334-9, pp. 593-603.
- Ye, Zheng. Zhan, Weidong. Zhou, Xiaolin. 2007. *The Semantic Processing of Syntactic Structure in Sentence Comprehension: An ERP Study*, Brain Research, Volume 1142, 20 April, 2007, pp. 135-145.
- Zhan, Weidong. 2004. 广义配价模式与汉语“把”字句的句法语义规则. 语言学论丛 No 29, pp. 334-368, Commercial Press 2004.