MT-Aided Analysis of Comprehension of English Verb Preposition Constructions

Ge Naisheng, Zhu Jiang, Wang Haifeng

Toshiba (China) Research and Development Center
5/F, Tower W2, Oriental Plaza, No.1, East Chang An Ave., Dong Cheng District
Beijing, 100738
{genaisheng, zhujiang, wanghaifeng}@rdc.toshiba.com.cn

Abstract

The paper presents a new method for analysis of natural language comprehension. In this paper, we narrow the analysis down to English verb preposition constructions, as they are very complicated not only in the composition but also in the meaning. Although we can make human or statistics investigation to study the comprehension of verb preposition constructions, human's incomplete memories and the irrelevance between human's comprehension and statistics survey cannot keep this problem into perspective. However, an MT system can both provide almost infinite memories and simulate a certain human comprehension. In this paper, by virtue of our E-C MT system and NIST evaluation metric, we want to know to what extent verb preposition constructions could trouble the comprehension process and which kind of the constructions could pose the biggest headache for learners. Having evaluated the scored translation results from our system by removing our system's rules with respect to all kinds of verb preposition constructions, we could figure out the difficult points of understanding different verb preposition constructions for English-learners.

1 Introduction

Verb particle construction is always defined to be the combination of a head verb and one or more obligatory particles, in the form of intransitive prepositions (e.g. hand in), adjectives (e.g. cut short) or verbs (e.g. let go) (Villanvicencio and Copestake, 2002a; Villavicencio and Copoestake, 2002b; Huddleston and Pullum, 2002). However, the verb preposition construction in this paper is different from the definition of verb particle construction. For the purpose of this paper, the verb preposition construction consists of head verb and any prepositions, or adverbial particles, or prepositional phrases. Thus it includes phrasal verb, prepositional verb, phrasal prepositional verb and free verb-preposition combination (see in section 2). These constructions are products of the recycling of finite preposition and verbs but render the language infinitely wealthy and expressive. Although we know a preposition usually indicates the temporal, spatial or logical relationship of its object to the rest of the sentence, many of us who have English as a Second Language still find preposition usages very frustrating. Why are they so difficult? On the one hand, prepositions are very tricky as they relate to culture specific views of how things are related in time and space, and in other logical frameworks. On the other hand, the meanings of some verb preposition constructions span the continuum between the nearly transparent compositional phrase and the fully opaque idiom (Don Blaheta and Mark Johnson, 2001).

Although we empirically know the importance of verb preposition constructions, we cannot quantitatively know how these constructions impeded our understanding. As a consequence, we want to examine these verb preposition constructions by virtue of Toshiba Machine Translation System (Amano, et al., 1989) and NIST evaluation metric

(Doddington, 2002) to explore the human comprehension difficulties presented by them.

We assume that as for studying a foreign language, we can use the rule-based MT system to simulate human comprehension process as a result of parallels between rule-based MT system development and human foreign language learning, that is, both of them necessarily accumulating language knowledge in almost similar way. In other words, human understanding a foreign language requires his stored language knowledge learned from the textbook or his daily foreign language experiences. And the human language knowledge precisely corresponds to a wide variety of rules for a rule-based MT system to translate a text.

From the methodical point of view, we could remove rules relating with constructions in question and evaluate the different translation results to investigate into the human comprehension of these constructions in a text with the help of NIST scoring tool.

The Toshiba Machine Translation System employs a semantic transfer approach. The grammar system used in this system is called Lexical Transition Network Grammar, which combines lexical grammars and a kind of augmented transition network (ATN) grammar. The parser implemented for this grammar analyses sentences syntactically with ATN and semantically with lexical grammars. The NIST scoring tool can automatically score the translation quality of a MT system by using n-gram co-occurrence statistics. The automatically generated scores have strong correlation with human assessments. Certainly, we can judge the translation results with human help rather than NIST scoring tool. Nevertheless, the point is that NIST scoring tool gives us more efficient and consistent judgments.

The remainder of this paper is organized as follows. In section 2, we describe the types of verb preposition constructions. In section 3 and section 4, we briefly introduce the Toshiba Machine Translation System and NIST evaluation metric. In section 5, we present our MT-aided analysis method. In section 6, we report the results and analysis of the experiment. In section 7, we present the conclusion and future work.

2 Verb preposition Constructions

As we have defined, verb preposition constructions can be classified into prepositional verb, phrasal verb, phrasal prepositional verb and free verb-preposition combination. In fact, prepositional verb, phrasal verb and phrasal prepositional verb can be altogether called multi-word verbs, which Quirk et al. (1985) defines as "a unit which behaves to some extent either lexically or syntactically as a single verb". Nevertheless, in free verb-preposition combination the head verb is only adverbially modified by prepositional phrase (PP) and its meaning has nothing to do with its PP complement.

2.1 Prepositional Verbs

Prepositional verbs are a group of multi-word verbs made from a verb plus prepositions and all prepositional verbs have direct objects:

1) intransitive verb + preposition
   Please **look after** the children.
   I have **fallen for** you in a big way.
2) transitive verb+ NP + preposition
   You have **talked** me **into** it.
   Rubber seals are fitted to **prevent** the gas **from** escaping.

In English, the preposition does not always come before its object; in certain kinds of sentence, it can come at the end of the clause:

What are you **talking about**?
What exactly are you **getting at**?

Besides, the prepositional verbs have the following characteristics:

1) commonly used prepositions: in, after, about, for, etc
2) never occur without object: talk about someone/something, etc
3) the direct object always follows the preposition

2.2 Phrasal Verbs

When a verb is used with an adverbial preposition the combination is called a phrasal verb. Phrasal verbs can be transitive or intransitive:

1) intransitive verb + particle
   The plane **took off** on time.
   Bullies often **back down** when you stand up to them.

2) transitive verb + particle

He **gave back** the book.

Resistance was finally **broken down**.

When a phrasal verb has a direct object, the two parts of the verb can usually be separated: the adverbial preposition can be put before or after the object.

We'll have to **put off the party**/**put the party off**.

However, when the object is pronoun, the adverbial preposition can only go after the object:

We'll have to **put it off**.

We'll have to **put off it**. (unacceptable)

To summarize the characteristics of phrasal verbs:

1) commonly used adverbs: up, down, off, on, etc.
2) when the object of the verb is present, it is placed either between the basic verb and the adverb or after the entire phrasal verb.

2.3 Phrasal Prepositional Verbs

Phrasal prepositional verbs consist of three parts: a base verb, an adverbial preposition and a preposition, such as, "get on with", "put up with", etc. These constructions look more complicated, but in fact, they are used in the same way as any other prepositional verbs.

1) Intransitive verb + particle + preposition

I won't **put up with** this noise.

Drop in on me some time.

2) transitive verb+ NP + particle +preposition

I will **take** you **up on** that.

He **gives** himself **over to** drinking, taking drugs and living in luxury.

The summarized characteristics of phrasal prepositional verbs are:

1) basic verb + particle + preposition (e.g. out of, up, for, etc)
2) always take a direct object
3) can not be separated by the object

2.4 Free Verb-Preposition Combination

Free verb-preposition combination, in contrast with multi-word verbs, is that verbs and prepositional phrases are defined separately in the sentence. The idiomatic prepositional phrases (e.g. at ease) are excluded in this paper.

They **talked over dinner**.

The boys are **playing in the garden**.

Dan came to the airport to **meet me at Los Angels**.

He ran **up the hill**.

Not include:

She had an ability to put people **at their ease**.

In truth, as to this free verb-preposition combination, we are more concerned about the usages of preposition as the flexibility of preposition usages always hinders our correct understanding of English.

At a word, verb preposition constructions are composed of two parts: multi-word verb and free verb-preposition combination. And multi-word verbs further fall into prepositional verbs, phrasal verbs and phrasal prepositional verbs. Some multi-word verbs are transparent (literal) as their meanings can be deduced by defining its individual part.

He **came in**.

Others are more or less opaque (idiomatic), the meanings of which are different from the meanings of the words taken separately.

He **gave up**.

As to the free verb-preposition combinations, they are always transparent.

She **called from the office**.

Of course, literal construction is easy to understand and idiomatic one is to some extent obscure. Now, it all comes down to the point to know the distributions and comprehension of these literal or idiomatic constructions in texts.

3 Brief Introduction of Toshiba E-C Machine Translation System

Toshiba E-C machine translation system is a rule-based E-C MT system. Its translation quality ranks among the above-average E-C MT systems available on the market. The basis of the system is semantic transfer, for which a grammatical model called 'Lexical Transition Network Grammar' was developed. This model is very powerful for semantic analysis, which is performed by lexical grammars attached to each item in the lexicon.

We take five steps in the translation flow: (1) dictionary and morphological analysis (2) syntactic analysis (3) semantic analysis (4) structural transfer (5) generation.

The syntactic analysis is purely syntactic, and syntactic rules have no semantic conditions. The syntactic analyzer is composed of two parts: syn-

tactic interpretation and structure construction. Syntactic interpretation consists of finding which syntactic rules may apply, thereby resolving lexical ambiguities. Structure construction re-applies these rules to build the syntactic representation.

The semantic analyzer constructs a semantic interpretation and simultaneously makes the conceptual structures for the target language. The system transfers directly to the target language and not via an interlingua for sake of efficiency. Based on the hypothesis: Meaning is lexical, in our system semantic rules are attached to words as lexical rules.

Structural transfer is another kind of transfer which is evoked by some features of syntactic structures. The meanings of lexical and structural transfer are different. Lexical transfer is a kind of converter form syntax to semantics while structural transfer is a converter from source language to target language and so it realizes a contrastive grammar.

Syntactic generation grammars decide the order of the nodes of the conceptual structure according to the syntactic role represented by the arcs and add/delete some words according to Chinese grammar.

4 NIST Evaluation Metric

We use the NIST scoring tool to automatically evaluate the translation results generated by our E-C MT system. The NIST evaluation metric, based on the n-gram co-occurrence evaluation method which was originally developed by IBM (Papineni, et al., 2002), serves as the primary evaluation measure for DARPA TIDES-sponsored MT research (Doddington, 2002).

The DARPA TIDES program is developing advanced technology for translingual information processing including information detection, extraction, summarization and translation. The Machine Translation Evaluation project is administered by NIST. In 2002, NIST conducted the first Machine Translation Evaluation (MT-02) in order to support MT research. MT-02 used n-gram co-occurrence evaluation to conduct automatic evaluation on Chinese-to-English and Arabic-to-English machine translation. For doing this, NIST developed an evaluation tool kit which could be downloaded directly from NIST website.

The n-gram co-occurrence evaluation scores a translation according to the n-grams that it shares with one or more reference translations. It is quick and language independent. Moreover, there is a strong correlation between the n-gram co-occurrence scores and human assessments of translation quality, which makes n-gram co-occurrence evaluation an efficient method for automatic evaluation.

To evaluate the translation quality of a given MT system with the NIST scoring tool, a set of reference translations is required except the source data and the MT system output translation. The source set, reference set and output translation set should be annotated according to predefined format. The scoring tool produces a translation quality score for the given MT system by comparing n-grams of the output translation with n-grams of the reference translations and counting the number of matches. Here n-gram means a sequence of n words. When we apply this tool to English-to-Chinese machine translation evaluation, we define n-gram as a sequence of n Chinese words. So Chinese word segmentation should be done before the evaluation.

5 MT-Aided Method

5.1 Why is MT-Aided Method

To detect the difficult points of comprehension of English verb preposition constructions, in this paper, we choose to employ our machine translation system and NIST scoring tool to simulate human comprehension process though we still can turn to other approaches such as pedagogical experiences from English teachers, statistics-aided analysis and subject questionnaires. If we could discuss the issue from the following aspects, however, we would infer that MT-aided method is the optimal choice.

First, the approach should accurately detect the verb preposition construction in a text. VPC[1] detection accuracy is the first factor considered.

Second, the approach should cover sufficient verb preposition constructions. Then we have VPC memory size.

Third, the approach should record the comprehension differences of detected verb preposition constructions. The third factor is the VPC comprehension difference.

[1] VPC is short for “verb preposition construction”

Forth, the approach should handle a large enough testing set to be objective. Testing set size comes into view.

Fifth, the approach should be executed with practical efforts. Efficiency is the last factor considered.

According to the above five factors, we have the following table.

	Detection accuracy	Memory size	Comprehension difference	Testing set size	Efficiency
PE[2]	+++	++	++	+	+
SA[3]	+	+++	-	+++	+
SQ[4]	+	+	+++	+	+
MA[5]	++	+++	+++	+++	+++

Table 1: Comparability of Candidate Approaches

We use the "+", "-" to evaluate these four approaches. Their meanings are as follows:

+++: best
++ : better
+ : good
- : bad

From the table, we can first eliminate the statistics-aided analysis because it cannot understand any constructions without human help, to saying nothing of the comprehension differences of constructions. Even with human help, the approach cannot guarantee the detection accuracy because of the absence of parser.

Pedagogical experiences from English teacher can have the best detection accuracy but even so, the teacher can not memorize all the verb preposition constructions and can not deal with a large enough testing set to make quantitative judgment. Moreover, the teacher is a proficient English speaker and he cannot truly reflect the confusion caused by verb preposition constructions and can only empirically sense the difficulties of verb preposition constructions. Therefore, his judgment is prone to be subjective.

Subject questionnaire only has the best factor of "comprehension differences" and its other several factors are not very good whereas it seems the second best choice because subjects' uncertain memorization and accuracy of verb preposition constructions could help differentiate the comprehension difficulties of VPC. However, the approach is faced with the same problem as the pedagogical experiences of English teacher: the limited testing set size and efficiency. To be objective, we need a lot of subjects with different English levels and present a reasonable but comprehensive questionnaire. The point is we cannot easily select subjects with English level as we wish. The task is huge and demands a lot of efforts.

The disadvantage of MT-aided analysis lies in the detection accuracy because its parsing results cannot be always correct. However, our system's parsing accuracy as regards VPC is almost 90.5% and the rules about VPC in our system correspond to almost all the VPC entries in several classic English dictionaries.

The advantages of MT-aided analysis consist in its almost infinite memory size of verb preposition constructions, as mentioned above, and the huge size of testing set it could handle. Another unique advantage is its automatic translation with certain rules absent and this enables us to select any "subjects" with different English levels at any moment. So this approach could get the best "comprehension differences". Furthermore, this "comprehension differences" is, to some extent, better than that of subject questionnaire because it is more controllable and feasible. Combined with automatic scoring tool, MT-aided analysis approach could easily help us to quantitatively analyze the comprehension difficulties of these verb preposition constructions in sentences.

5.2 Description of MT-Aided Method

The key points of MT-Aided method are reliable parsing results, sufficient rules concerned, objective testing set and believable scoring tool.

In our system, the parsing accuracy of whole sentence is 83% and parsing accuracy of verb preposition construction is 90.5%. Up to now, we have over 10,000 rules with respect to verb preposition constructions, covering most of entries related with verb preposition constructions in several classic dictionaries. NIST scoring tool is a prevalent utility for automatic MT evaluation. The testing set is a corpus containing 1,000 sentences randomly achieved from our bilingual corpus. Therefore, we believe we have the adequate condi-

[2] PE: pedagogical experiences
[3] SA: statistic-aided analysis
[4] SQ: subject questionnaires
[5] MA: MT-aided analysis

tions for the MT-aided analysis of comprehension of verb preposition constructions.

Before experiment, we investigate into the distribution of verb preposition constructions in the testing set. The occurrence of each kind of VPC is counted.

Our experiment begins with translating the testing set with no rules removed. The translation is scored by NIST tool as a standard to be compared with other results.

Next, we remove the rules of above four types respectively and score the translation results one by one.

Finally, we remove the rules of the multi-word verbs (including phrasal, prepositional verb, and phrasal prepositional verb) and score.

In our system, removing rules of some constructions does not mean that the system would fail to translate the sentence containing the constructions. Instead, the sentence is translated by other rules.

For example, considering the sentence "He is looking after his wife", if the rule of "look after" is removed, the system would translate "look" and "after" separately rather than no translation. In other words, the system would select the default rule of a word if it could not find other rules of the word. In our system, the default rule of a word is always the most frequent usage of a word and this is as well the basis for a learner to guess a new phrase. So this enhances the idea that the translating way of our MT system is much similar to human thinking way when faced with an unfamiliar phrase.

6 Experimental Results and Analysis

6.1 Experimental Results

First we look into the distribution of verb preposition constructions in the testing set. The distribution is based on the occurrences of each construction.

Second we respectively score the testing set after removing the rules about each construction. And we also show the percentage of rules of each construction in the whole rules of verb preposition constructions.

Then we look at the score fall of each construction compared with the standard score achieved without removing rules. Meanwhile, the percentage of score fall is achieved by dividing score fall of whole verb preposition constructions by the score fall of each construction.

So we have the following table.

	RNP	OCRP	Scores	SF	SFP	Gap
FVC	3.4%	59.0%	9.4744	0.3096	41.9%	-17.1%
PHVC	19.4%	10.8%	9.6817	0.1023	13.9%	3.1%
PPVC	7.4%	2.8%	9.7617	0.0223	3.1%	0.2%
PRVC	69.7%	27.4%	9.5155	0.2685	36.6%	9.2
MVC	96.6%	41.0%	9.3959	0.3881	52.9%	11.9%
VPC	100%	100%	9.0510	0.7330	100%	0%
Standard			9.7840			

Table 2: Evaluations of verb preposition constructions

FVC: Free Verb-preposition Combination
PHVC: PHrasal Verb Construction
PPVC: Phrasal Prepositional Verb Construction
PRVC: PRepositional Verb Construction
MVC: Multi-word Verb Construction
VPC: Verb Preposition Construction
Standard: removing no rules

RNP: Rules Number Percentage of each verb preposition construction.

$$\text{RNP}=\frac{\textit{rules number of each construction}}{\textit{rules number of VPC}}\times 100\%$$

OCRP: OcCuRrences Percentage of each construction in the testing set

$$\text{OCRP}=\frac{each\,of\ construction}{vpc}\times 100\%$$

Scores: after removing rules of each construction

SF: Score Fall of each construction compared with the standard score

SF =scores of standard – scores of each construction

SFP: Score Fall Percentage of each VPC in the score fall of VPC.

$$\text{SFP}=\frac{SF\ of\ each\ \ construction}{SF\ of\ VPC}\times 100\%$$

Gap is the difference between SFP and OCRP

Gap=SFP – OCRP

Both SFP and OCRP are about the ratios between each kind of VPC and all VPC in the testing set. The gap comparison between them is helpful to evaluate the comprehension of each construction.

6.2 Experimental Analysis

The experiment shows that the scores of different verb preposition constructions do not differ greatly from the standard. This is not a surprise in that verb preposition constructions are only part of the grammatical construction of a whole sentence. As a result, just removing the rules of verb preposition constructions cannot change the translation results totally. But, it is enough for us to evaluate the comprehension of different verb preposition constructions.

First we compare the two constructions: free verb-preposition combinations (FVC) and multi-word verbs construction (MVC).

We assume that if the effect of all constructions on comprehension is similar, SFP of one construction should roughly correspond to its OCRP. But we can find the wide gap between OCRP and SFP. As to the gap, we believe that it indicates the trend of increasing or decreasing role of one construction in the comprehension of verb preposition constructions. Now we just consider the gap to be the trend of role of a construction, and its exact meaning will be studied later in our future work.

Obviously the bigger SFP represents the stronger effect on the comprehension. As a result, if SFP of one construction is higher than its OCRP, we deduce its increasing role in the comprehension and vice versa.

According to Table 2, we learn that SFP (59.0%) of MVC is higher than its OCRP (41%). The gap is 11.9%. However, SFP (41.9%) of FVC is much lower than its OCRP (52.9%) of MVC. The gap is –17.1%. Hence, we infer that MVC has the stronger effect on the comprehension than FVC based on the above assumption. As a matter of fact, the inference is consistent with the idea that the meanings of MVC are either transparent or opaque and the meanings of PVC are always transparent.

In table 2, we learn that the rules of free verb-preposition combinations constitute 3.4% among the total rules of verb preposition constructions, and those of multi-word verb constitute 96.6%. These data about free verb-preposition combination indicate that we could only learn a small number of usages to understand free verb-preposition combination, however, learning multi-word verb requires understanding much more times usages of FVC.

In addition, we see from the table that the OCRP of FVC and MVC is respectively 59.0% and 41.0%. It means we will encounter more FVC than MVC when we read an article.

From the above discussions it is natural for us to come to the following deductions:

1. Compared to multi-word verb constructions, free verb-preposition combination appears more frequently in the English text.
2. It is much easier to understand free verb-preposition combination than multi-word verb constructions.

From the above deductions, we could further infer that when we are faced with an unfamiliar verb preposition construction, we naturally take it as a free verb-preposition combination. As we invariably literally understand free verb-preposition combination, sometimes we would be right if coincidentally the meanings of unknown verb preposition constructions were transparent (literal), and sometimes we would be at fault if the meanings were opaque (idiomatic).

Now the question boils down to the examination of MVC comprehension since we always un-

derstand their meanings from either literal or idiomatic point of view. We are supposed to compare the effect of each construction of MVC, namely, PHVC, PRVC and PPVC on the comprehension of verb preposition construction.

We notice that contrary to the FVC whose meaning is always transparent, all the SFP of PHVC, PPVC, PRVC is higher than their OCRP, quantitatively proving the fact once again that their meanings could not be always literally understood. As we have discussed, the gap between SFP and OCRP suggests the trend of role of each construction in the comprehension. The biggest gap of PRVC, which is 9.2% compared to the 3.1% of PHVC and 0.2% of PPVC, reflects the greatest effect of PRVC on comprehension among them.

In addition, PRVC has the largest RNP (69.7%), OCRP (27.4%), and SFP (36.6%) among the MVC.

Therefore, we could learn that in the MVC constructions, PRVC not only have the highest frequency of all but also have the strongest influence on the comprehension. This strongest influence is also supported by the fact that this construction usually has the pattern: verb + noun + preposition + noun, which cover relatively more words in a sentence, thus maybe having greater impact.

Summarily, we learn free verb-preposition combination has the widest distribution in the English text, but it has absolutely minimum rules in the system. Therefore, it is not a surprise that when we meet with verb preposition constructions, the thought pattern about free verb-preposition combination lures us into literally understanding verbs and prepositions because it is easy to think this way and it more often than not pays off. Consequently, we sometimes are not only blind to the some more or less idiomatic multi-word verbs but also unaware of the study for them. The inevitable confusing point is that it seems to us that we know all the words in an article, but we just cannot catch it.

7 Conclusion

In our experiment we quantitatively prove that the multi-word verbs have the strongest influence on the comprehension of verb preposition constructions and PRVC is the biggest headache in multi-word verbs. At same time, their comprehension is often influenced by the free verb-preposition combination. If we could focus on the study of multi-word verb, especially PRVC, we could improve our English comprehension more quickly.

In the future, we will continue to improve our E-C MT system. We will also take into consideration the analysis of comprehension of more English constructions. And the further research on correlation between human and MT-aided analyses of natural language comprehension will be made. Meantime, adopting our method in more applications, such as the analysis of the development process of the rule-based MT system, is also in our mind.

Acknowledgements

We would like to thank Guo Yuqing, Satoshi Kinoshita, Akira Kumano, Tatsuya Izuha, Yumiko Yoshimura, and other colleagues from Toshiba R&D Center for their contribution on developing the E-C MT system.

References

Shin-ya Amano, Hideki Hirakawa, Hirosysu Nogami, Akira Kumano. 1989. The Toshiba Machine Translation system. *Future Computing System Volume 2, Number 3* pages 227-246.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leach, and Jan Svartvik. 1985. A comprehensive grammar of the English language. *Longman Press*.

D. Blaheta and M. Johnson. 2001. Unsupervised learning of multi-word verbs. *In Proceedings of the ACL workshop on Collocations*, pages 54-60, *Toulouse, France.*

Alline Villavicencio and Ann Copestake. 2002a. Phrasal verbs and the LingGo-ERG. *LingGo Working Paper No.2002-01*.

Aline Villavicencio and Ann Copestake. 2002b. Verb-particle constructions in a computational grammar. In *Proc. of the 9th International Conference on Head-Driven Phrase Structure Grammar (HPSG-2002)* pages 357-371

Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. *In Proceedings of ACL-02*, pages 311-318.

G. Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. *In Proceeding of the Second International Conference on Human Language Technology*, pages 138-145