

# Research on Verb Semantic - syntactic knowledge base in Russian - Chinese Machine Translation

Li Yan

College of Foreign Languages and Cultures,  
Xiamen University, Xiamen, China  
liyan\_386@163.com

**Abstract**—In this study, the research results at home and abroad on the basis of semantics, through the establishment of the Russian verb semantic pattern of combing and significance patterns, to analyze the performance of syntactic distribution, description of the Semantic - syntactic patterns, and as the basis for constructing Machine Translation "verbal semantic - syntactic pattern library", "fixed sentence library" and "Variable library" to complete the carding three databases to achieve a knowledge integration. This study not only can expand the research category of semantics theory, but also can promote the further development of the Russian-Chinese machine translation theory.

**Keywords**—Russian-Chinese machine translation; semantic - syntactic information database; semantic description

## I. INTRODUCTION

Machine translation (MT) is realized from a natural language text (source language) into another natural language text (target language) translation by computer. Essential attributes common between different languages is semantic, common semantic understanding of the original premise is also guaranteed to be accurate expression<sup>[1]</sup>. Gui Du domestic famous linguist once said: "Through research semantic syntax, principles, structural constraints of meaning, the meaning of syntactic constraints, etc., for the development of computational linguistics and machine translation and other language learning application provides a theoretical and practical basis<sup>[2]</sup>. We can say that the semantic analysis restricts the specific process of machine translation, semantic recognition and re-encoding of the source language and the target language is the key to improving the quality of translation studies of this subject is from the Linguistics. departure, based on a combination of semantic relations between verb and a noun, the establishment of Machine Translation Russian verbs semantic - syntactic information database.

## II. REVIEW OF CURRENT RESEARCH

### A. Research Status Review of European countries

Research on machine translation Western countries began in the 1940s, the early machine translation system mainly uses direct translation method, that word conversion mainly focus on the word form. This approach lacks support linguistic translation theory, the quality is not high, is not satisfactory.

1966 American Language Processing Advisory Committee automatically (ALPAC) for machine translation research made a negative conclusion, claiming Machine Translation encountered insurmountable "semantic barriers", affected many countries Machine Translation a low ebb. But ALPAC speech allows researchers are beginning to recognize the important role in the semantic analysis of machine translation, and made a series of research results. Machine Translation broader application of a semantic analysis technology is "Case Grammar ".The theory proposed sentence structure level semantics and syntax associated with it trying to find out the deep structure into surface structure of the law. But the " Case Grammar " In theory, there is a clear thematic shortcomings is that it free the dominant verb, relying on logic or semantic intuition vaguely discuss and find generally have a deep semantic grid. Since the 1980s, based on indirect methods of translation has made considerable progress, conversion analysis from simple syntax level of development to the syntax and semantics of a combination of deeper, such as Kaplan R. and Bresnan J.'s "Lexical Functional Grammar", due to the dominance of too much emphasis on grammar, and can't solve the puzzle machine translation semantic barriers. Overall, the machine translation in morphology and syntax, it has already been proven, through technical and general method for the test, but in semantic analysis technology has not been well studied.

### B. Russia Status Review

Research on semantic theory, Russian scholars have made a more results. One of the most prominent is to Апресян Ю. Д. and Мельчук И. А. represented by the Moscow semantic school. In 1974, the famous Russian scholar Мельчук И.А. initiated and created a new system of grammatical theory - "means  $\Leftrightarrow$  Text" model (Модель «Смысл $\Leftrightarrow$ Текст», МСТ), its computational linguistics, machine translation applied research to provide a set of syntactic rules from the semantic system. After 40 years of development, MCT academic influence in the international computing linguistics expanding in Russia, Canada, France, Germany, Japan and other countries have launched MCT model design based machine translation system, or try in the native language of the target language codification bilingual dictionaries, etc., has made achievements in many applications, a representative of the "Russian Comments with dictionary" (TKC). But generally speaking, TKC there are still many problems, give it to those who:

- ① did not establish a unique and systematic conversion

mechanism.

② in describing aspects of the verb, its interpretation limited to "single-level semantic structure" did not make the induction and summed up the performance of semantic pattern in different languages or categories of verbs.

③ from the isolated, dispersed and lack of systematic analysis, word for word, the necessity of thematic, optionally, etc. did not make a strict distinction.

Further, the study is for the English machine translation system, the development of Russian-Chinese translation system is relatively slow. Russia in science and technology machine translation, Russian-English, Russian-French accuracy rate up to 80%, while the accuracy of Russian-Chinese translation of less than 50% (Fu Xingshang, 2009). In short, the formal semantics of the problem is very complex, although in recent decades it made attempts, but has not yet made substantive progress.

### C. Review of the status of domestic

Our research in machine translation started earlier, but the degree of attention and research level Russian-Chinese machine translation research far and Chinese - English translation studies, also behind the Western countries. Representatives are Zhang Jiahua, Xue Enkui, Peng Yuhai, Yi Mianzhu, Fu Xingshang, Xu Hancheng, Li Xiangdong, Feng Zhiwei like. They draw on the Moscow semantic school advocates of "Words Function" and "Comments dictionary" and so on, to study the word sense disambiguation problem<sup>[3]</sup>. Representative works are Zhang Jiahua "Theory and Application" Glossary function "" (2002), and the monograph "Russian Contemporary Semantics" (2003). Fu Xingshang also "Glossary function" applications in Russian-Chinese machine translation is discussed<sup>[4]</sup>.

In summary, the Russian-Chinese machine translation, the inadequacy of semantic analysis are mainly: when it comes to the different meanings of the verb is given the number of valence and form, failed to go deep into the semantic distinction about the meanings of form content, this leads to differences in the various meanings of the verb can't be an objective description, it is difficult to resolve ambiguity (polysemy, structural ambiguity semantic ambiguity, etc.) machine translation between multiple meanings, and thus triggering conflicts choice. Therefore, the focus on formal semantics of language research today, with the syntactic features language component, to make an objective description on similarities and differences in semantics which is very important.

## III. MACHINE TRANSLATION OF VERB - SYNTAX DESCRIPTION SYSTEM

Based on the meaning to the form of functional grammar perspective, given the need for verbal semantic first category and second category description, then the syntax for the content of this operation (sense) materials, verification of the Semantic - syntactic description of the objective, scientific and feasible degrees. Language structure, take the following steps deep into the hierarchy of verbal semantic - syntactic restriction of this

nature to a whole category of verbal semantic make a complete and detailed description.

### A. Verb combinations relationships (syntactic distribution aspects):

1) The number of valence verbs and syntactic functions. To find the same number of thematic and valence forms of the verb in the verb semantic class, which is the most important performance of Verb Semantic and formal classification basis.

2) Thematic times verb semantic features. Observe and analyze the same person verb core structure in terms of thematic similarities and differences of views semantic features, which is a unique form of testing level.

3) Verb "co-occurrence compatible language" (temporary increases). Further analysis of the possible collocation verb, it is primarily a prime verb outside of the core structure, which generally include the verb "optionally price" and "Optional price."

### B. Verbs aggregation relationship (conversion feature aspects form):

1) Synonymous conversion. Words do not change the sentence refers to the syntax and semantics of the composition and the relationship between the elements or in the same sentence for achieving semantic conversion. Syntax and semantics of the verb form of the distribution relationship could not be distinguished from, you can find differences in a variety of ways synonymous conversion<sup>[5]</sup>. Moreover, even if there are different combinations of performance, but also can make further distinctions in such conversion characteristics. For example:

$$N_1^a V f N_4^b \leftrightarrow N_1^a v f < \text{на} > N_6^b$$

(Снег давит ветви. ↔ Снег давит на ветви.)

This mode allows the conversion sentences behavior object to obtain additional "orientation" meaning, and become a "super-theta."

2) Reverse translation. Words do not change the composition of the sentence means, around the same root word or grammar derived reverse reverse word for thematic form - syntax against the bits, and so on through a main object of the verb form of the conversion swap roles. Different semantic categories inverse change in the form of the verb also has a profound reflection. Abundant use of inverse exchange can unfold the same way as the corresponding language in the form portrayed on deep semantic mechanism verbs. For example:

$$N_1^a V f < \text{на} > N_4^b \leftrightarrow N_1^a v f < \text{в} > N_4^a$$

(Сапог не идет на ногу. ↔ Нога не идет в сапог.)

This is changing the verb "prefix type" reverse change. The semantics of justice is strong, mainly seen in "wearable" type verbs.

Here we have more дать verb, for example, include semantic model and its discriminative formal semantic description language release in Table 1:

Table 1. Verb дать Semantic - syntactic pattern analysis

Semantic model	the dominant mode	the thematic sub-semantic features compatible co-occurrence features	the thematic sub-semantic features compatible co-occurrence features	reverse conversion feature
CC1 give, dedicated	$N_1^a V E N_4^b N_3^c$ дать отдать посвятить	$N^a$ :чел. $N^b$ :предм., одуш. $N^c$ :чел.	сп.,	$N_1^a V E N_4^b N_3^c \leftrightarrow N_1^b V E_{c,a} A_5^a$
CC2 grants, awards	$N_1^a V E N_4^b N_3^c$ дать вручить присвоить присудить	$N^a$ :чел. $N^b$ :предм. $N^c$ :чел.	сп., оц. прич.	
	$N_1^a V E N_4^b N_5^c$ наградить	$N^a$ :чел. $N^b$ :чел. $N^c$ :предм.		
CC3 supply, provide	$N_1^a V E N_4^b N_3^c$ дать предоставить	$N^a$ :чел. $N^b$ :предм. $N^c$ :чел.		
	$N_1^a V E N_4^b N_5^c$ обеспечить	$N^a$ :чел. $N^b$ :чел. $N^c$ :предм., одуш.		
	$N_1^a V E N_4^b N_5^c$ снабдить	$N^a$ :чел. $N^b$ :чел. $N^c$ :предм.		

CC – семантическая схема abbreviation, CC1 namely semantic mode 1, and so on.

"Thematic sub-semantic features" and "compatible with the co-occurrence" column in the abbreviated code index:

чел.	People	предм.	Object name	оц.	Product Reviews
прич.	The reason	одуш.	Living thing	сн.	The way

#### IV. RUSSIAN AND CHINESE VERBAL SEMANTIC - SYNTACTIC KNOWLEDGE BASE

##### A. The use of previous studies, extract verb semantics selected Russian and Chinese bilingual corpus - syntactic patterns.

① Real amount of extraction. The Russian verbs, verbs with matching related phrases and sentences extracted into Russian semantic - syntactic pattern of the real amount; the real amount of Russian semantic pattern corresponding to the amount of the Chinese translation of extracts into a real Chinese semantic model.

Real amount means the semantic model irreplaceable natural language characters<sup>[6]</sup>, using the "R" represents. In verbal semantic model, the verbs and prepositions and verbs matched as actual amount and other terms. With "V" represents a verb, "O" indicates the amount of other entities. "R" constituted by the "V" or combination "V" and "O" of. In some specific natural language, "R" Available corresponding specific language vocabulary to express. For example:

Я завидую ему. 我羡慕他。

N [people] завидовать N [people]; N [people] 嫉妒 N [people].

Он играет в футбол. 他在踢足球。

N [people] играть в N [ball]; N [people] 踢 N [ball]。

Among them, the first sentence in Russian and Chinese "завидовать" real amount "R", the V components. "Завидовать" for the "V" portion. The second sentence "играть в" real amount "R", the "V" and "O" combination thereof. Wherein, "играть" to "V" portion, " в " to "O" portion.

② Extraction variables. The same position in Russian syntax, semantics same type of partially extracted for the same semantic - syntactic patterns of variable; the variable semantic pattern corresponding Russian English translation extract into a variable Chinese semantic model.

Variables are part of the semantics of words affect verbal semantic model can be replaced. Variables with an "X". Variables can be replaced by one or more semantic types in a specific semantic model. For example:

Он спит. 他睡觉。

N[people] спать; N[people] 睡觉。

Анна знает русский язык. 安娜懂俄语。

N[people] знать N[language]; N[people] 懂 N[language].

Рабочие строят дом из камня. 工人用石料盖房。

N[people] строить N[thing] из N[material]; N[people] 用 N[material] 盖 N[thing].

Among them, the first sentence of the Russian and Chinese semantic variable from one type N [people] constitution; the second sentence of the variables from the two semantic type N [people] and N [language] composed; and the third sentence variables for three semantic type N [people] N [people], N [matter] and N [material] composed.

*B. On the basis of the establishment of Russian - Chinese Verbs knowledge base.*

These include:

1) *Verbal semantic - syntactic pattern library.* Russian main storage verb semantic - syntactic patterns, Chinese semantic - syntactic patterns, semantic - syntactic patterns the number of semantic types, variables, and variables, Russian and Chinese sentences sentence, which holds artificially summary of verbal semantic - syntactic patterns infos to provide a basis for the semantic pattern matching to identify and verbs Translation work.

2) *Fixed sentence library.* By a number of convention contain verbs, indivisible Russian English translation of the

*sentence and the corresponding configuration sets. Fixed sentence stored in the library sentence for example as follows:*

患难见真情 Друзья познаются в беде

祸不单行 Беда не проходит одна

一言为定 Сказано-сделано.

3) *Variable Library.* Semantic types and from a variety of Russian and Chinese in particular represents a collection of corresponding configuration. For example:

Semantic Type N [direction] corresponds to the Russian and Chinese specifically expressed as: восток 东方; запад 西方; север 北方; юг 南方; северо-восток 东北方; северо-запад 西北方; юго-восток 东南方; юго-запад 西南方。

Semantic Type N [season] corresponds to the Russian - Chinese specifically expressed as: весна 春季; лето 夏季; осень 秋季; зима 冬季。

Verbal semantic model library structure and library structure variables are shown in Table 2 and Table 3.

Table 2. The structure of verb semantic pattern base

Name	Data Types	Field Size	Remark
Number	Automatic Number	Long integer	Verbal semantic model number
The number of variables	Digital	156	Variable number of records
Variable Type	Text	45	Variable type record
Russian semantic pattern	Text	156	Record Russian verb semantic pattern
Chinese-defined modes	Text	156	Chinese Semantic verb record mode
Semantic pattern type	Text	45	Semantic type recording verbal semantic model
Russian sentence	Text	156	Recording double statement against the Russian sentence
Chinese Sentence	Text	156	Record double the statement of the Chinese Sentence

Table 3. The structure of variable base

Name	Data Types	Field Size	Remark
Number	Digital	Long integer	Variable Number word
Russian word	Text	45	Russian record variable word form
Chinese translation	Text	45	Record variable forms of Chinese words
Semantic Types	Text	45	Record variable word semantic types

Here we have the verb *дать* synonymous, for example, include semantic model library (Table 4).

Table 4. Example of V semantic Pattern base

Verb	the number of variables and types	Chinese semantic model	semantic model Russian	Chinese sentences	Russian sentences
дать	[1] $N_1^a$ [人] [2] $N_4^b$ [人、物] [3] $N_3^c$ [人]	[1]给[3][2]	[1]дать[2][3]	妈妈给孩子们水果。	Мама дала детям фрукты.
	[1] $N_1^a$ [人] [2] $N_4^b$ [职务] [3] $N_3^c$ [人]	[1]授予[3][2]	[1]дать[2][3]	他授予我勋章。	Он дал мне орден.

вручить	[1] $N_1^a$ [人] [2] $N_4^b$ [奖章、奖品、物品] [3] $N_3^c$ [人]	[1]授予[3][2]	[1]вручить[2][3]	市长授予他勋章。	Мэр вручил ему орден.
наградить	[1] $N_1^a$ [人] [2] $N_4^b$ [人] [3] $N_5^c$ [奖章、奖品]	[1]奖励[2][3]	[1]наградить[2][3]	老师奖励孩子书。	Учитель наградил детей книгами.
присвоить	[1] $N_1^a$ [人] [2] $N_4^b$ [荣誉、称号、学位、军衔] [3] $N_3^c$ [人]	[1]授予[3][2]	[1]присвоить[2][3]	市长授予他称号。	Мэр присвоил ему звание.
отдать	[1] $N_1^a$ [人] [2] $N_4^b$ [物品] [3] $N_3^c$ [人]	[1]把[2]还给[3]	[1]отдать[2][3]	他把书还给我。	Он отдал мне книгу.
посвятить	[1] $N_1^a$ [人] [2] $N_4^b$ [物品] [3] $N_3^c$ [人]	[1]把[2]献给[3]	[1]посвятить[2][3]	他把诗歌献给母亲。	Он посвятил стихи матери.
обеспечить	[1] $N_1^a$ [人] [2] $N_4^b$ [人] [3] $N_5^c$ [物品、人力]	[1]给[2]提供[3]	[1]обеспечить[2][3]	我们给您提供专家。	Мы обеспечиваем вас специалистами.
снабдить	[1] $N_1^a$ [人] [2] $N_4^b$ [人] [3] $N_5^c$ [物品]	[1]给[2]提供[3]	[1]снабдить[2][3]	我们给您提供材料。	Мы снабжаем вас материалами.
предоставить	[1] $N_1^a$ [人] [2] $N_4^b$ [物品、权利] [3] $N_3^c$ [人]	[1]给[3]提供[2]	[1]предоставить[2][3]	我们给您提供资金。	Мы предоставляет ему средства

## VI. CONCLUSION

Development Trend of language requirements on the basis of the theory of syntax and semantics of the value system of highly formalized, the establishment of a set of language connotation logic operable. Based on statistics, the classification of other concepts, to move closer to computational linguistics syntax has become an important part of contemporary research, however, positive lexicogrammar theory needs to be under the guidance of semantic syntax check mode. In other words, "syntax with semantics can't stand alone as the language evolution of standards and scale, they need to combine as a whole"<sup>[7]</sup>. Thus, the subject of semantic - syntactic theory to build a "verbal semantic - syntactic pattern library," "fixed sentence library" and "Variable library" to complete the carding three databases to achieve a knowledge integration.

Currently, we have to extract the meaning of entries more than 100 Russian verb from the dictionary while including Chinese interpretation and so on. Examples from our Russian National Corpus (НКРЯ), ТКС, "Russian-Chinese Dictionary Explanation" (four volumes) and "with Russian and Chinese dictionary." Machine translation is a difficult and complicated task, Translation Studies verb to reach a satisfactory result, but also the need for verbal semantic model more accurate and scientific division, and continuous improvement of existing semantic model libraries, build a more complete verbal

semantic model library.

## ACKNOWLEDGMENTS

This research topic is sponsored by Fujian Social Sciences Planning Program (No. 2013C082), National Social Science Fund (No. 13BY043), and Xiamen University Fund for Prospering Philosophy and Social Science (No. 0650-Y07200)

## REFERENCES

- [1]. Shu Dingfang, Modern Semantics Shanghai: Shanghai Education Press, 2000.
- [2]. Du Guizhi, "Meaning of the principles of integrated language syntax Restrictions," China Russian Teaching, 2010 1, pp. 6-11.
- [3]. Zhang Jiahua, Syntax • Semantics • Pragmatics - Modern Russian Studies, Heilongjiang People's Publishing House, 2000.
- [4]. Fu Xingshang, Xu Hancheng, Yi Mianzhu, Li Xiangdong, Russia computational linguistics and machine translation, Beijing: Chinese Press, 2009.
- [5]. Peng Yuhai, Semantico - Syntax Integral Descriptions of The Russian Language, Heilongjiang People's Publishing House, 2008.
- [6]. Zhang Yao, Research on Chinese Translation of English Verbs in English-Chinese Machine Translation . Dalian: Dalian University of Technology, 2010.
- [7]. Степанов Ю. С, Имена, предикаты, предложения (Семиологическая грамматика). Москва. 1981.