A Study on the Formalization of English Subjunctive Mood

Xiaowei Guan[*]

School of Foreign Languages, Dalian University of Technology, Dalian, China
Email: angel_gxw@yahoo.com.cn

Abstract—**One of the main problems that affects the quality of machine translation is how to express the knowledge of language in precision. Subjunctive mood is a very common language phenomenon in English. From the perspective of E-C machine translation, and based on the theory of Semantic Element (SE) in Unified Linguistics, the paper discusses the specific formalization methods for each type of the English sentences with subjunctive mood and their Chinese translations.**

Index Terms—**subjunctive mood, formalization, SE, SER**

I. INTRODUCTION

The history of machine translation is more than a half century long, but its quality is still poor and the machine translation systems still haven't reached the practical phase. The key lies in linguistics, that is, the knowledge of language cannot be expressed in precision, although many famous linguists have made many important achievements from different angles of study.

Subjunctive mood is a very common language phenomenon in Indo-European languages, which is a verb mood typically used in subordinate clauses to express a wish, emotion, possibility, judgment, opinion, necessity, or action that has not yet occurred. The subjunctive in Modern English is easily distinguished in a variety of contexts in which the present subjunctive is used where the sense is past tense. So the tense and aspect are often combined with the mood in English. In Chinese, there isn't the concept of subjunctive mood. But the subjunctive meaning is expressed by some words and even implied in sentences and situations in Chinese.

In linguistics, subjunctive mood is a very important grammar, which has been studied by many scholars in the aspects of classification, syntax, usage, mood, translation in both written and oral English (Zhang, 2002; Chalker, 1995; Zhang, 1998). In NLP, until now subjunctive mood has not been studied as a special and individual topic.

This paper offers an account of one key problem of the precision of natural language: the formalization of English sentences with Subjunctive Mood and their corresponding Chinese sentences. In section 2, we discuss the theory and concept for the formalization based on Unified Linguistics. Section 3 shows the specific formalization methods for each type of the English sentences with subjunctive mood.

II. THE THEORY AND CONCEPT FOR THE FORMALIZATION

A. The Theory of SE and SER

In 1980's, Institution of Computing Technology (ICT) of the Chinese Academy of Science (CAS) proposed a new concept (Gao, 1989), i.e., semantic element representation (SER) with variables and without variables. The concept of Semantic Element and Semantic Language was discussed in detail in the paper of Gao (2003; 2009).

According to Unified Linguistics (Gao, 2009) , different language can be translated into each other and people speaking different language can communicate with each other because there are words, phrases and sentences with the same meaning. Natural languages need to be described in a unified way from the perspective of semantics. Semantic of a sentence is called SS. An element to express a semantic meaning in an SS is called Semantic Element (SE). Semantic language (SL) consists of all SEs, including all SSs. The representation of an SE in a natural language-I, such as English, Chinese…, is called the Representation of Semantic Element in Language-I (SER_I). Semantic of SER is SE. Any natural language can be regarded as a representation of semantic language. The translation between two languages (I, J) is regarded as a transformation between two representations.

For example, the Chinese sentence is "李先生是教授 *Li xiansheng shi jiaoshou* (Mr. Li is a professor.)". The four SEs in this sentence are 李(Li)，先生($X_{surname}$) (Mr. (X))，教授 (professor)，是 $_{title}(X_{people}, Y_{title})$ ($Is_{title}(X_{people}, Y_{title})$). "X" and "Y" are two parameters. SE is an abstract concept and form. Actually, the above mentioned examples of SEs are only some remembrance forms to represent SE by using Chinese characters or English words. We can use "1" to substitute the SE "Li", or use "$4(X_{people}, Y_{title})$" or "$4(N_{people}, N_{title})$" to substitute "是 $_{title}(X_{people}, Y_{title})$". The subscript

denotes the semantic category of the word. An SER can reflect the semantics and syntax relations of the word in a phrase or a sentence.

B. Other Relevant Concepts about the Formalization

English sentences with subjunctive mood are formalized based on the theory of SE and SER. SER is the specific representation of an SE in a certain language. SER_c denotes Chinese SER; SER_e denotes English SER.

SE consists of two kinds in its structure: SE without parameter (WP-SE) and SE with parameter (P-SE). SER consist of three kinds in its structure: simple SER (S-SER), general SER (G-SER), and abstract SER (A-SER). S-SER corresponds to WP-SE, G-SER and A-SER correspond to P-SE. G-SER is composed of fix parts (words, characters, word clusters in a language) and parameters. But A-SER only consists of parameters and the operation among them. A-SER corresponds with abstract SE (A-SE) (Gao, 2009).

The parameter is formed by two parts, in which one part is a symbol like "N, A, V, S" to represent noun, adjective, verb and sentence respectively, the other part is the subscript denoting the semantic category of the word. For example, "N_{person}" denotes nouns with the semantic category of person. Words with the same semantic category can be substituted by each other in an SER. Not only words but also phrases, clauses and sentences can be formalized into parameters.

The SE and SER theory has been applied into the formalization of Chinese classifier-noun phrases, English prepositional phrases and sentences (Guan, 2009). This theory has also been widely applied into other fields, such as machine translation, handwritten character recognition, Internet information monitoring and post-processing technique for speech and character recognition (Li, 2006; Gao, 2004; Gao, 2005).

III. THE FORMALIZATION OF ENGLISH SENTENCES WITH SUBJUNCTIVE MOOD

A. The Types of English Sentences with Subjunctive Mood

Unreal conditional sentence is the most common subjunctive mood to express hypothesis. It consists of two kinds--one is contrary to the present reality, the other is contrary to the past reality. For example: "If I were you, I should buy it. 如果我是你 *ruguo wo shi ni*, 我就买了它 *wo jiu maile ta*", and "If you had got up earlier, you could have caught the train. 如果你早一点起床 *ruguo ni zaoyiidan qichuang*, 就会赶上火车的 *jiuhui ganshang huiche de*".

The object subordinate clause after verbs like "wish, suggest, order, insist, propose, etc" should use subjunctive mood to express a wish, request, suggestion and command. For example: "He insisted that all of us should be there on time. 他坚持要我们大家按时去那儿 *ta jianchiyao women dajia anshi qu naer*", and "I prefer that he should end his speech. 我更希望他结束演讲 *wo geng xiwang ta jieshu yanjiang*".

Subjunctive mood in subject subordinate clause exists in the sentence pattern "It is A that…", in which A is the adjective with the meaning of importance and necessity. For example: "It's necessary that we should have a walk now. 我们有必要出去散散步 *women youbiyao chuqu sansanbu*".

The appositive subordinate clause after nouns like "wish, suggestion, proposal, advice, etc" should also use subjunctive mood to express a wish, requirement, suggestion and order. For example: "It is my suggestion that the students be sent to Tibet.我建议把那些学生派到西藏去 *wo jianyi ba naxie xuesheng paidao Xizang qu*".

There are other phenomena of subjunctive mood in both simple and compound sentences to express a doubt, supposition, wish, etc. For example: "If only he had seen me!要是他看见我就好了 *yaoshi ta kanjian wo jiuhaole*!", "He would rather go by bus than by train. 我宁可坐汽车也不坐火车 *wo ningke zuo qiche yebu zuo huoche*", and "But for the rain, we should have had a pleasant journey. 若不是下雨，我们那次旅行就惬意了 *ruobushi xiayu, women naci lüxing jiu qieyi le*", etc."

B. The Formalization of English Sentences with Subjunctive Mood

For an English sentence with subjunctive mood, the parts that express the subjunctive meaning should be extracted into the fix structure and other parts of the sentence should be extracted into parameters to form P-SE.

Since some of the subjunctive mood is related to the tense and aspect of verbs, we propose a new concept "Style". Style refers to the mood, voice, tense, aspect, comparison, number, gender, case, modal, etc. Different languages may have different Style which can be realized by transformation. The introduction of "style transformation" aims at the precise unification of different languages. Style is a kind of A-SE. The combination of Style and A-SE (SE with Style) is also an A-SE. The combination of Style and the parameter of SE (parameter with Style) is also an A-SE.

Case 1: For the simple sentence "If only he had seen me!要是他看见我就好了 *yaoshi ta kanjian wo jiuhaole*!", there are the following style transformations for part of this sentence: SER_e(seen(he′, me′))=<S>=he see me. (without subjunctive mood/tense style)→SER_e((seen, *past perfect*) (he′, me′))= <S>=he had seen me (with past perfect tense style). Here, SER_I(X) denotes the representation of SE X in Language-I, (seen, *past perfect*) shows SE (seen(he′, me′)) with style. The SE and SER for the whole sentence are: SS=if $_{only}$(S)=要$_{是就好了}$(S); SER_e=if only S!; SER_c=要是 S 就好了! The part "if only" in this sentence denotes the subjunctive mood which should be extracted as the fix structure in the SER_e, and the corresponding Chinese translation part will also be extracted as the fix structure in the SER_c. The other words in the sentence should be extracted as the parameters in both SER_e and SER_c.

Case 2: In the simple sentence with modal verb "He could have taken a taxi. 他原本可以坐出租车去 *ta yuanben keyi zuo chuzuche qu*", the modal verb "could" and auxiliary verb "have" should be extracted as the fix part in the SER. The SE and SER of this sentence are: SS=could$_{have}$(N$_{person1}$, S-Z)=原本可以(N$_{person}$, S-Z); S-Z=(taken, V_{ed}) (N$_{vehicle}$)= 坐$_{...去}$(N$_{vehicle}$); SER$_e$= N$_{person1}$could have S-Z; SER$_c$= N$_{person}$原本可以 S-Z. Here, Z denotes the subject of a sentence; S-Z denotes a sentence without subject. Here, V_{ed} denotes the verb of past participle.

Case 3: For the complex sentence "If I were you, I should buy it. 如果我是你 *ruguo wo shi ni*, 我就买它 *wo jiu maile ta*", since the verb "be" can only be the form of "were" in subjunctive mood, "were" should be extracted as the fix part in the SER. The SE and SER of this sentence are: SS=if (S$_1$, S$_2$)=如果(S$_1$, S$_2$); S$_1$= were(N$_{person1}$, N$_{person2}$)=是 (N$_{person1}$, N$_{person2}$); S$_2$=should(N$_{person1}$, S-Z)=就(N$_{person1}$, S-Z); S-Z=buy(N$_{object}$); SER$_e$=if S$_1$, S$_2$; SER$_c$=如果 S$_1$, N$_{person1}$ 就 S-Z.

For another complex sentence which has the same sentence structure "If I'd heard the whole story, I would have taken another measure. 如果我听到了事情的全部经过 *ruguo wo tingdaole shiqingde quanbu jingguo*，我就会采取另外一种做法 *wo jiuhui caiqu lingwaiyizhong zuofa*", we can use another method to formalize this sentence. There exist the following style transformations for the first clause of this sentence: SER$_e$(heard(I′, story′))=<S$_1$>=I hear the story (without subjunctive mood/tense style)→SER$_e$((heard, *past perfect*) (I′, story′))= <S$_1$>=I had heard the story (with past perfect tense style). Here, (heard, *past perfect*) shows SE (heard(I′, story′)) with style. The SE and SER for the whole sentence are: SS=if (S$_1$, S$_2$)=如果(S$_1$, S$_2$); S$_2$= would have(N$_{person}$, V$_{ed}$, N)=就会(N$_{person1}$, V, N); SER$_e$=if S$_1$, N$_{person}$ would have V$_{ed}$ N; SER$_c$=如果 S$_1$, N$_{person}$ 就会 V N. Or we can represent the sentence as: SS=if (S$_1$, S$_2$)=如果(S$_1$, S$_2$); S$_2$= would have (N$_{person}$, S-Z)=就会(N$_{person}$, S-Z); S-Z=(taken, V_{ed}) (N) = 采取(N); SER$_e$=if S$_1$, S$_2$; SER$_c$=如果 S$_1$, N$_{person}$ 就 S-Z. The concept "Style" is shown again in formalize this sentence, just as that in case 1. The "style transformation" denotes a kind of tense transformation in this kind of translation of English subjunctive mood, especially in C-E translation.

Case 4: In the complex sentence "I prefer that he should end his speech.我更希望他结束演讲 *wo geng xiwang ta jieshu yanjiang*", the word "prefer" and "should" both indicate the subjunctive mood, so these two words should be the fix parts in the SER, and other words should be extracted as parameters. The SE and SER of this sentence are: SS=prefer$_{that...should}$(N$_{person1}$, N$_{person2}$, S-Z)=更希望(N$_{person}$, S); S-Z=end(N$_{event}$)=结束(N$_{event}$); SER$_e$= N$_{person1}$ prefer N$_{person2}$ S-Z; SER$_c$= N$_{person}$更希望 S. In this sentence, "prefer that" and "should" are both the symbol of subjunctive mood, so they should both be extracted as the fix parts in the SER.

Case 5: In the complex sentence "It's necessary that we should have a walk now. 我们有必要出去散散步 *women youbiyao chuqu sansanbu*" and "It is my suggestion that the students be sent to Tibet.我建议把那些学生派到西藏去 *wo jianyi ba naxie xuesheng paidao Xizang qu*", the words "necessary", "suggestion" and "should" all indicate the subjunctive mood. So just as case 3, they should be the fix parts in the SER. Moreover, the sentence structure "It is…that…" should also be as the fix part. SE and SER of the two sentences are: SS=It $_{is\,necessary\,that...[should]}$(N$_{person}$, S-Z)= 有$_{必要}$(N$_{person}$, S-Z); SER$_e$= It's necessary that N$_{person}$ [should] S-Z; SER$_c$= N$_{person}$ 有必要 S-Z；SS=It $_{is...suggestion}$ $_{that...[should]}$(A$_{possessive}$, N$_{person}$, S-Z)=建议(N$_{person}$, S); SER$_e$= It is A$_{possessive}$ suggestion that N$_{person}$ [should] S-Z; SER$_c$= N$_{person}$ 建议 S. Here, the symbol [ ] indicates the word inside can be omitted.

Case 6: The sentence "He would rather go by bus than by train. 我宁可坐汽车也不坐火车 *wo ningke zuo qiche yebu zuo huoche*" is a subjunctive mood sentence with a set structure. For this kind of sentence, we extract the set structure as the fix part in SER. So the SE and SER of this sentence are: SS=would $_{rather...than}$(N$_{person}$, S-Z, B)=宁可$_{...也不}$ (N$_{person}$, S-Z, S-Z); B=by(N$_{vehicle}$); SER$_e$=N$_{person}$ would rather S-Z than B; SER$_c$= N$_{person}$ 宁可 S-Z 也不 S-Z. Here, B indicates prepositional phrase.

For different kinds of subjunctive mood sentences, we need to use different formalization method, in which the fix parts and parameters in SER are extracted according to the sentence features. Above all, SER is extracted correspondingly from bi-lingual sentences for the sake of English-Chinese translation.

C.   The Construction of SER-base of English Subjunctive Mood

For each type of the English subjunctive mood and their formalizations, we construct the SER-base of subjunctive mood (SM-SERB), which includes the English sentences of subjunctive mood marked by adverbs (e.g. if, rather, although, etc), verbs (e.g. prefer, suggest, etc), adjectives (advisable, necessary, etc), modal verbs (e.g. could, should, etc) and subjunctive mood sentences with set structures. The English sentences with subjunctive mood, their Chinese translations, style, SER$_e$ and SER$_c$ should be stored in SM-SERB using the software Access.

The SER-base will be applied into some specific language studies. For example, the SER-base can be used as a corpus for the study on the translation of English subjunctive mood both in translation teaching and study; the SER-base can provide the rules for the machine translation of English subjunctive mood into Chinese; moreover the base can partly help us to correctly select the tense and aspect of the verbs while translating the Chinese sentence into English.

IV.   CONCLUSION

This paper focuses on solving one of the key problems of describing language phenomenon precisely in MT—the formalization of English sentences with subjunctive mood and their Chinese translation. Specific formalization methods are proposed based on the theory of SE, SER and Style in Unified Linguistics. The study in the paper will beneficial to the improvement of the quality of subjunctive mood translation in MT. More efforts should be made to put this method into practice.

REFERENCES

[1] Zhang, D. Z. (2002). A Practical English Grammar. Beijing: Foreign English Teaching and Research Press. (in Chinese)
[2] Li, L., Gao, Q. S. (2006). A Method for Improving Offline Handwritten Character Recognition Based on Semantic Analysis. *Computer Engineering and Applications 42(6)*, 10-12. (in Chinese)
[3] Gao, Q. S., Chen, Z. X. & Li, T. Q. (1989). The Principle of Human-like Machine Translation. *Computer Research & Development, 26*, 1-7. (in Chinese)
[4] Gao, Q. S., Hu, Y., Li, L. & Gao, X. Y. (2003). Semantic Language and Multi-Language MT Approach Based on SL. *J. Comput. Sci & Technol, 18*, 848-852.
[5] Gao, Q. S., & Gao, X. Y. (2004). The Method, Device and Auto-translation System of Improving the Accuracy of Speech and Character Recognition. 200410062566. (in Chinese)
[6] Gao, Q. S. & Gao, X. Y. (2009). Foundations of Unified Linguistics. Beijing: Science Press. (in Chinese)
[7] Chalker, S. (1995). Dictionary of English Grammar. Oxford: Oxford University Press. (in Chinese)
[8] Guan, X. W. (2009). Study on Several Key Problems in the Machine Translation System Based on Semantic Language. Ph.D. dissertation, Dalian University of Technology. (in Chinese)
[9] Gao, X. Y., Gao, Q. S., Hu, Y. & Li, L. (2005). High Speed Multi-language Machine Translation Based on Pruning on the Tree of Representations of Semantic Elements. *Journal of Software, 16(11)*, 1909-1919.
[10] Zhang, Z. B. (1998). New English Grammar Course. Shanghai: Shanghai Foreign Language Education Press. (in Chinese)

Xiaowei Guan was born in Shenyang, China in 1979. She received her PH.D. degree in computer application from Dalian University of Technology, China in 2009.

She is currently a lecturer in the School of Foreign Languages, Dalian University of Technology, Dalian, China. Her research interests include machine translation and natural language processing.