A CLASSIFICATION OF CHINESE VERBS FOR LANGUAGE PARSING

Li-li Chang*, Keh-jiann Chen**, Chu-ren Huang***, Ching-chun Hsieh*,**

* Computing Center, Academia Sinica, Taipei, Taiwan, R.O.C.
** Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.
*** Institute of History and Pilology, Academia Sinica, Taipei, Taiwan, R.O.C.

February 10, 1988

EXTENDED SUMMARY

1. Verb Classification and Language Parsing

For parsing, verb classification is a vital part of syntax, because verbs are heads of sentences. An appropriate classification of verbs can help a lot in constructing the whole structure of a sentence, thus in grasping the semantic meaning of the sentence.

In most linguistic theories, verbs are classified according to their syntactic properties, such as transitivity, argument types, etc. Such syntactic classifications provide a way of defining different phrase structure rules and sentence patterns of each different class of verbs. However, classification based purely on syntactic terms can not provide too much semantic information. Some linguistic theories, such as Lexical-functional Grammar and Case Grammar, also provide mechanisms to explicitly represent functional or semantic role assignment to constituents. We adopt such a mechanism — case frame, named after Case Grammar. Due to the fact that positions of arguments of verbs in a Chinese sentence is highly flexible, case frames for Chinese verbs also serve the purpose of identifying the syntactic phenomenon, such as topicalization, relativization, object preposing , etc. To serve all these purposes, the steps we follow in classifying verbs are:

(1) what is the case frame of this verb? (2) what are the sentence patterns associated with this verb? and (3) what are other useful syntactic or semantic features?

In theory, it is not necessary to identify the classes which each verb belongs to. Once information are provided according to the three steps, verbs sharing the same grammatical information forms a natural class. Thus, we are more or less following the current theory of feature based categorization. A category is a set of elements with the same feature value set. In such a way, we avoid the problem of multiple classification, but fulfill the purpose of sentence parsing.

In practice, to give a clearer picture and to reduce the storage requirement, we still provide the higher level classifications for each verb. Since 'case frame' is a semantic structure of the verb, verbs of the same subset classified according to their meanings must have similar case frames and thus the redundant data can be reduced. By this method, we classify the verbs in Mandarin into the following three main classes, i.e. active verbs, which have six subtypes, stative verbs, which has also six subtypes, and classificatory verbs, as shown in (1). Every subtype has its own pattern of case frame, but details may differ from verb to verb.

(1) active verbs: intransitives
                    transitives
                    ditransitives
                    verbs with sentential complement
                    pivotal verbs
                    verbs with verbal complement
     stative verbs: intransitives
                    transitives
                    ditransitives
                    verbs with sentential complement
                    pivotal verbs
                    verb with verbal complement
     classificatory verbs

2. Case Frame

An active verb implies the notin of 'action', which must be performed by an 'agent', being the syntactic subject in typical Mandarin sentences. The class of stative verb is used to describe an entity whose case type is termed in our system a 'patient'. Syntactically this entity is always the subject. That means an active verb must have an argument playing as an 'agent', while a stative verb has a 'patient', and both arguments are also the subject of sentences. We further subdivide the verbs according to the number and the types of other arguments, i.e. post-verbal constituents. Clear description of the case frames and case restrictions of each subtype will be given in the full paper.

Each argument of a verb has its own particular semantic restriction, and this informaion can help us to identify the function of the arguments. Thus, based on our case frames, the following sentences should be successfully parsed, though in the following sentence pairs, underlined phrases of identical syntactic category (NP) and identical linear order play different grammatical roles.

(2) a. <u>我</u> 看過了。
'I have seen (it).'
b. <u>那部電影</u> 看過了。
'That movie, I have seen it.'

(3) a. <u>你</u> <u>那件工作</u> 完成了 沒有？
'Have you finished the work?'
b. <u>那件工作</u> <u>你</u> 完成了 沒有？
'The work, have you finished?'

(4) a. <u>本公司</u> 供應 <u>鮮奶</u>。
'Our company supplies fresh milk.'
b. <u>這些鮮奶</u> 供應 <u>各超級市場</u>。
'Those fresh milk will be supplied for every supermarket.'

(5) a. 本公司 供應 美國 各市場。

     'Our company supplies every market in American.'

   b. 本公司 供應 美國 各種火腿。

     'Our company supplies every kind of beacon in American.'

     'Our company supplies American with every kind of beacon.'

(6) a. 我 介紹 張三 大兒子 一位漂亮的小姐。

     'I introduce one beautiful girl to Jangsan's eldest son.'

   b. 我 介紹 張三 我的辦公室 一位漂亮的小姐。

     'I introduce one beautiful girl in my office to Jangsan.'

(7) a. 張三 送 玩具 汽車。

     'Jangsan sent one toy car (to someone).'

   b. 張三 送 李四 汽車。

     'Jangsan sent Lisz one car.'

(8) a. 張三 送 李四 玩具 汽車。

     'Jangsan sent lisz one toy car.'

   b. 張三 送 李四 爸爸 汽車。

     'Jangsan sent Lisz's father one car.'

3. A Set of Sentence Patterns

To parse more efficiently, in addition to its case frame, we give every verb the syntactic information, i.e. the sentence patterns a verb may occur. Such information can also help us to subclassify the verbs. For example, we can seperate 'pseudo transitives', which are semantically transitive but syntactically intransitive, from 'transitives', which are both semantically and syntactically transitive. Beside transitivity, there are other syntactic properties which can help to classify verbs, such as the positions arguments may appear, the prepositions an argument may be marked with, etc.

Number of sentence patterns are rather large, therefore we have to determine the domain of our encoding. There are basically two principles:

first, we consider only simple sentences; secondly, sentence patterns which are appliable to all the verbs in a class are set as defaults. In the paper, we will give a detailed discussion about the types of the patterns concerning these two principles.

4. Conclusion

Our system provides three kinds of information about a verb: (1) its required case types with their semantic restrictions, (2) its sentence patterns, and (3) its semantic and syntactic features. By the first and part of the third kind of information, we classify the verbs into three main classes and thirteen subtypes. By the second information, we can further subdivide those subtypes. For example, the subset of 'transitive verbs', active or stative, can be divided into two types: the transitives and the pseudo transitives.

Our classification of verb has four advantages for language parsing. First, it can construct the sentence structures much more efficiently no matter how the constituents are ordered, such as the instances in (2)-(8). Secondly, the semantic features such as the distinction between active verbs and stative verbs can help to solve many parsing problems. For example, if a superficially coordinate sentence is composed of one active phrase followed by a stative one, as in (9), the first clause is usually an adverbial phrase indicating the temporal dimension, while the second one is the main clause.

(9)張三看書喜歡坐在床上。
'While reading books, he likes sitting on the bed.'

Third, some semantic properties can predict a fixed syntactic representation. For example, the argument which implies the notion 'event' instantiates as a syntactic sentence, regardless of whether the argument is pre- or post-verbal. Fourth, by connecting case restrictions cited in all case frames, we can construct a conceptual structure. Such structure will be a hierarchical semantic structure of nouns. On the same par, our classification of verbs is

- 5 -

a hierarchical semantic structure of the verbs. Together they can contribute
to the construction a complete hierarchical structure of knowledge, which will
be another important model in a parsing system.