The NLP-Oriented Empty Categories

Aiping Tu

Center for Study of Language and Information, Wuhan University, Wuhan, China 430072

tuaiping81@163.com

Keywords: NLP-oriented, empty categories, nominal, verbal, prepositional.

Abstract. Empty Categories are expressions which are base-generated in the deep structure and have no overt phonetic form in the surface structure. As for computers, it is difficult to make clear where and what the empty categories are. Therefore, revealing the empty categories is essential to the nature language processing (NLP). Chinese does not only have nominal empty categories, but also verbal and prepositional empty categories. This article proves the existence of these two non-nominal empty categories in Chinese, and further points out where and what the empty categories are.

Introduction

With the emergence of computer and the gradual popularization of internet, the nature language processing (abbreviated as *NLP*) has become more and more important. Among which, the understanding of empty categories (abbreviated as *ECs*) is the most difficult one. According to the economic principles, in some cases, an expression may not present all constituents of deep structure, which reflects the basic semantic relations, and has no overt phonetic form in the surface structure. According to the economic principles, in some cases, an expression may not present all constituents of deep structure, which reflects the basic semantic relations, and has no overt phonetic form in the surface structure. As a result, empty categories which are base-generated in the deep structure but exist covertly in the surface structure can often be seen. Consider below.

(1) a. seem John like apple → John$_i$ seems [t_i to like apple]

b. John see what → What$_i$ did John see t_i?

c. John try John to win the race → John$_i$ try *PROi* to win the race.

d. Zhangsan$_i$ chi pingguo$_j$? → *pro$_i$* chi *pro$_j$*

　Zhangsan$_i$ eat apple$_j$? → *pro$_i$* eat *pro$_j$*.

　'Does Zhangsan eat apple?' →'He eats it.'

The *t, PRO* and *pro* [1,2](Shen calls it *Omission e* [3]), which exist in deep structure but have no overt phonetic forms in surface structure, are called empty categories. It is generally acknowledged that ECs have been studied with a main focus on nominal empty categories (abbreviated as *NEC*), which occupy the positions of the nominal constituents in the syntactic structure [4-8]. Verbal and prepositional empty categories (abbreviated as *VEC* and *PEC* respectively), however, have been less discussed. The NLP-oriented language researching calls for the revealing of the interface between syntax and semantics. This article reveals the existence of VECs and PECs in Chinese, and further makes clear where and what the ECs are.

The Existence of the VECs in Chinese

It is well known that NECs are obligatory arguments which have no overt phonetic form in surface structure, and can be infered by the argument structures of core verbs. The theta-criterion requires a one-to-one correspondence between an argument and its theta-role. Vice versa, it can be concluded that the NPs with theta-roles must have their corresponding theta-role assigners.

In English, an argument must have its corresponding theta-role assigner which has an overt phonetic form, or it is ungrammatical. Consider (2).

(2) a. John is student. → * John *VEC* student.

b. John has become an undergraduate. → * John *VEC* an undergraduate.

c. John has three apples. → * John *VEC* three apples.

But the same cases in Chinese are grammatical. In Chinese NECs and VECs are also base-generated in deep structures, but have no overt phonetic form in the surface structure. Consider below.

(3) a. Zhangsan shi xuesheng. → Zhangsan *VEC* xuesheng (, Lisi *VEC* Laoshi).
Zhangsan is student → Zhangsan *VEC* student (, Lisi *VEC* Laoshi).
'Zhangsan is student.'

b. Zhangsan biancheng daxuesheng le. → Zhangsan *VEC* daxuesheng le.
Zhangsan become undergraduate SF→ Zhangsan *VEC* undergraduate SF
'Zhangsan has become an undergraduate.'

c. Zhangsan you san ge pingguo. →Zhangsan *VEC* san ge pingguo.
Zhangsan has three CL apple →Zhangsan *VEC* three CL apple
'Zhangsan has three apples.'

VP-internal subject hypothesis points out that all arguments get their theta-roles within the VP, and the specifier position of IP is only the landing-site of the subject after its movement [9]. The nominal constituents like *Zhangsan* and *san ge pingguo* in (3c) must be arguments with theta-roles, otherwise, they cannot have their semantical interpretation in Logical Form (LF).

However, the NPs in sentences like *Zhangsan* and *xuesheng* in (3a) and *Zhangsan* and *daxuesheng* in (3b) are not arguments in the traditional sense and the verbs in them are linking verbs, they have meaning, but they cannot be a predicate by themselves, and can be predicates only when they form predicates together with their complement. And the two NPs in this type of sentence are not of the same type: the former is the entity and the latter acts as the predicate together with linking verb.

The Existence of the PECs in Chinese

Except for verb, preposition is another word class that can assign accusatives. Although verbs assign the Case of NPs in object positions, especially those who plays a role like agent, experiencer, beneficiary, patient, theme, the Case of NPs that donate roles such as location, instrument, method, reason is assigned by prepositions. In English, just like the verbs that assign accusative cannot be covert, the prepositions cannot be covert either. See below.

(4) a. John had lunch *in* hall.→*John had lunch (*in* PEC) hall.
b. John lives *on* the car.→*John lives (*on* PEC) the car.

Just like the cases of covert verbs, prepositions in Chinese can also present without overt phonetic forms. The typical ones are those in the unconventional object sentences. All unconventional object sentences have their reversibility, as the unconventional objects can be preceded by prepositions and then form PPs which are composed of prepositions and their objects [10]. Consider (5).

(5) a. Zhangsan chi shitang. →Zhangsan *zai* shitang chi.
Zhangsan eat mess hall → Zhangsan in mess hall eat
'Zhangsan has lunch *in* mess hall.'

b. Zhangsan chi zhe liang che. →Zhangsan *kao* zhe liang che chi.
Zhangsan eat this car → Zhangsan depends this CL car eat
'Zhangsan lives *on* the car.'

The unconventional objects can have their semantical explanation by using the concept of PEC. According to the uniform theta-role assignment hypothesis / UTAH) [11], verbs assign the theta role of *patient / object* to the object after the verb. For example, the verbs like *chi* can only assign accusative Case to the *pro* (which is *patient / object*). Therefore, the unconventional objects in (5) do not have their semantical explanation and the accusative Cases, these unconventional objects are not the obligatory argument of *chi*, and must get accusative Case from the corresponding prepositions. It is reasonable to infer that the prepositions which should assign accusative Case to unconventional objects present with the covert form.

Where and What the ECs are

Native speakers can understand the ECs in a sentence at the frst sight, but as for foreign learners or computers, it is difficult to decide where and what the ECs are. Therefore, revealing the ECs is essential to the nature language processing.

When the VEC is a copula, the second NP indicates the property or type of the first one. The interpretation of this type of sentence is not dependent on its context. The main copula is *be* in English and *shi* in Chinese [12,13]. See (6a). When the two NPs in a sentence are equivalent, the function of *shi* is equivalent to the linking verb *equal* in English, and the covert '*shi*' in the sentences can be replaced by the symbol '='. Consider (6b).

(6)a. Ta (shi$_{VEC}$) huang toufa.
  She (is$_{VEC}$) yellow hair
  'She has blond hair.'

 b. Jintian (shi$_{VEC}$) 2012nian yi yue yi ri. → Jintian = 2012nian 1yue 1ri.
  Today (is$_{VEC}$) 2012 January 1st → Jintian = 2012, January 1st.
  Today is January 1st, 2012.   → Today = January 1st, 2012.

Another type of VEC is other linking verbs except copula. The VECs in existential sentences indicating the presence, figurative sentences, as well as the 'NP *le*', do not require the context to help understand, and can be interpreted as the linking verbs *exist*, *look like* and *become* respectively, see (7).

(7)a. Wo jia (cunzai$_{VEC}$) liang ge haizi.
  I family (exist$_{VEC}$) two CL child
  'There are two children in my family.'

 b. Haizi de lian (xiang$_{VEC}$) huar shide.
  Children –'s faces (look like$_{VEC}$) flower seem
  'Children's faces look like flowers.'

 c. Ta (chengwei$_{VEC}$) daxuesheng le.
  He (become$_{VEC}$) undergraduate SF
  'He becomes a college student.'

Chinese VECs include not only covert linking verbs, but also notional verbs. Generally, they can be interpreted in isolation, i.e. be inferred by the most possible theta-roles of its arguments. But if put it into the specific context, just like the *pro* in pro-drop language, the empty notional verb in (3c) may vary according to the specific context. It can be other verbs, such as *chi, mai* or *reng*, and can be filled out according to the relevant contexts. Compare the two sentences in (8).

(8)a. Zhangsan (you$_{VEC}$) san ge pingguo.
  Zhangsan (ha$_{VEC}$) three CL apples
  'Zhangsan has three apples.'

 b. Zhangsan (chi$_{VEC}$) san ge pingguo.
  Zhangsan (eat$_{VEC}$) three CL apple
  'Zhangsan has eaten three apples.'

Compare with VECs, the semantic range of PECs are rather narrow. For example, the Case assigners of location are prepositions such as *zai, yu,* the Case assigners of instrumental are prepositions such as *yong, yi.* If the ECs in the unconventional object sentences are filled out, including conventional objects of *chi*, which exist with the form of *pro,* and the accusatives designer of unconventional object, which exist with the form of *PEC*, the following sentences can be composed.

(9) a. Zhangsan *zai*$_{PEC}$ shitang chi mou shiwu$_{pro}$.
  Zhangsan in $_{PEC}$ mess hall eat some foods$_{pro}$.
  'Zhangsan eats some foods in mess hall.'

 b. Zhangsan *kao*$_{PEC}$ zhe liang che chi fan$_{pro}$. → Zhangsan chi zhe liang che.
  Zhangsan on$_{PEC}$ this CL car live → zhangsan live on this car
  'Zhangsan lives *on* this car.'

From the analyses above, such a conclusion can be drawn: if a two place verb has only one overt obligory argument in surface structure, another must be covert, i.e, NEC. The same is true for the NECs which are obligatory arguments of one place or three place verbs. Only an argument with

theta-role can get its semantical interpretation, and the VECs and PECs can be inferred by the theta-roles of the arguments.

Concluding Remarks

In brief, revealing the nature of language is essential to the nature language processing, and NLP-oriented language research requires comprehensive observations to the overt and covert forms of syntactic constituents. ECs generally indicate nominal expressions such as empty nouns (or pronouns). However, except for NECs, Chinese also has VECs and PECs. Arguments with theta-roles and overt NPs with Cases require the existence of verbs or prepositions. These verbs and prepositions can be either overt or covert. When covert, they are termed as VECs and PECs.

Just like the NECs, VECs and PECs are also base-generated in the deep structure and have covert expressions in the surface structure, and can be inferred from the overt expressions. The NECs can be infered by the core verbs, and the VECs and PECs can be inferred by the theta-roles of arguments.

Acknowledgements

References

[1] Noam Chomsky. Lectures on Government and Binding. Dordrecht: Foris (1981)

[2] P Barbosa. Pro-drop and Theories of pro in the Minimalist Program[J] Language and Linguistics Compass No5/8:(2011)

[3]Yang Shen. A Study on Empty Category in Modern Chinese. Jinan: Shandong Education Press (1994)

[4]C-T James Huang. Logical Relations in Chinese and the Theory of Grammar. Ph.D. dissertation, MIT (1982)

[5] Liejiong Xu The Nature of Null Objects. Chinese.Journal of Chinese Language and Computing No16: (2006)

[6]Y-H Audrey Li. The Theory of Empty Categories and the Identif ication of Empty Categories in Chinese. Language Science, No1:(2007)

[7]Shukhan Ng. Processing Chinese Empty Categories. Ph.D. dissertation, The City University of New York (2009)

[8] Heyou Zhang & Sze-wing Tang. Licensing of Empty Categories and the Syntactic Variation of Copular Topic Sentences in Mandarin and Cantonese. Language Science, No 1:(2011)

[9] Dominique Sportiche. The Theory of Floating Quantifiers and Its Corollaries. Linguistic Inquiry, No 19:(1988)

[10]Shengli Feng. Light Verb Movement in Modern and Classical Chinese. Language Sciences, No 1:(2005)

[11]M. Baker. Incorporation. Chicago: University of Chicago Press (1988)

[12]Noam Chomsky. Knowledge of Language: Its Nature, origin and Use. New York: Praeger (1986)

[13]Bruce Horton. What are copula verbs?. Cognitive Linguistics in the Redwoods. ed. by Eugene H. Casad, New York:Mouton de Gruyter (1995)