Annotating the Predicate-Argument Structure of Chinese Nominalizations

Nianwen Xue

University of Colorado
The Center for Spoken Language
Boulder, CO 80309-0594
Nianwen.Xue@colorado.edu

Abstract

This paper describes the Chinese NomBank Project, the goal of which is to annotate the predicate-argument structure of nominalized predicates in Chinese. The Chinese Nombank extends the general framework of the English and Chinese Proposition Banks to the annotation of nominalized predicates and adds a layer of semantic annotation to the Chinese Treebank. We first outline the scope of the work by discussing the markability of the nominalized predicates and their arguments. We then attempt to provide a categorization of the distribution of the arguments of nominalized predicates. We also discuss the relevance of the event/result distinction to the annotation of nominalized predicates and the phenomenon of incorporation. Finally we discuss some cross-linguistic differences between English and Chinese.

1. Introduction

This paper describes the Chinese NomBank Project, the goal of which is to annotate the predicate-argument structure of nominalized verbs in Chinese. The Chinese NomBank extends the general annotation framework of the English Proposition Bank (Palmer et al., 2005), the Chinese Proposition Bank (CPB) (Xue and Palmer, 2003), and the English Nombank (Meyers et al., 2004) to the annotation of nominalized verbs in Chinese. Like the CPB project, the Chinese NomBank adds a layer of semantic annotation to the Chinese TreeBank (CTB) (Xue et al., 2005), which is an on-going project that currently has a corpus of 500 thousand words. The Chinese Nombank annotates two types of elements that are associated with the nominalized predicate: argument-like elements that are expected of this predicate, and adjunct-like elements that modifies this predicate. Arguments are assigned numbered labels (prefixed by ARG, e.g., ARG0...ARGn) while adjuncts receive a functional tag (e.g., TMP for temporal, LOC for locative, MNR for manner). A predicate generally has no more than six arguments and the complete list of functional tags for adjuncts and their descriptions can be found in the annotation guidelines of this project.

The Chinese NomBank also adds a coarse-grained sense tag to the predicate. The senses of a predicate, formally called framesets, are motivated by the argument structure of this predicate and are thus an integral part of the predicate-argument structure annotation. Sense disambiguation is performed only when different senses of a predicate require different sets of arguments. These senses are the same senses defined for the corresponding verbs in the Chinese Proposition Bank, but typically only a subset of the verb senses are realized in their nominalized forms. The example in Table 1 illustrates the Chinese NomBank annotation of two nominalized predicates “发展/development” and “规划/planning”, and the Nombank annotations added to the parse tree are in bold. The frameset identifiers for both predicates are f1. Of the four expected arguments for “发展/development”, ARG0 the cause or agent, ARG1 the theme, ARG2 the initial state and ARG3 the end state or goal, only ARG1 is realized and it is 两岸关系 (“cross-Straits relations”). “发展/development” also has a modifier labeled ARGM-TMP, 合后 (“hereafter”). The other nominalized predicate in this example, “规划/planning”, has two possible arguments, ARG0 the planner and ARG1 the thing that’s being planned. ARG0 is “海峡/Strait两岸/两岸” and ARG1 is realized as a prepositional phrase “regarding 合后两岸关系/cross-Straits relations”.

While this general annotation framework is the same as the English and Chinese Proposition Banks, in this paper we will focus on a few issues that are unique to the annotation of nominalized predicates. The rest of paper is organized as follows. In Section 2, we will first outline the scope of our work and discuss the markability and non-markability of the nominalized predicates and their arguments. Certain nouns in Chinese are not true nominalizations even though there exists a verb that shares the same morphological form with it. In addition, not all modifiers of a nominalized predicate can be considered as its arguments or adjuncts. Some modifiers only occur with the nominal form of a predicate but never with its corresponding verb form. In this case they are not considered to be arguments of the predicate. In Section 3, we show that the ways in which arguments are realized are very different for verbs and their nominalizations. The arguments of a nominalized predicate are often realized within a noun phrase. However they can occur outside of a noun phrase when the predicate occurs with a support verb. We then discuss the relevance (or irrelevance) of the event/result distinction to the annotation of nominalized predicates and the phenomenon of argument incorporation in Section 14. In Section 5, we discuss some cross-linguistic differences between Chinese and English. Finally, Section 6. summarizes this paper.

2. Scope of the work

The goal of this project is to annotate the argument structure of the nominalized verbs (or deverbal nouns). We start by sifting through all nouns that shares the morphological form with the nominalized predicate. It is not always the case that the nominal form of the predicate is the same as the nominalized predicate. For example in Table 1, the nominalized predicate 发展/development also has a modifier labeled ARGM-TMP, 合后 (“hereafter”). The other nominalized predicate in this example, 规划/planning, has two possible arguments, ARG0 the planner and ARG1 the thing that’s being planned. ARG0 is 海峡/Strait两岸/两岸” and ARG1 is realized as a prepositional phrase regarding 合后两岸关系/cross-Straits relations”.

¡Unless otherwise noted, by morphological form we mean the way a word is written, not by how it is pronounced. This is admit-
have similar argument structures as their verbal counterparts. Not all nouns that have a similar verbal form is a nominalization. For example, 行政 can be used as both a noun (“executive authority”) or a verb (“exercise executive authority”). While the nominal form and verbal forms are clearly related, the do not share the same arguments. The verbal form is a predicate that requires an agent, an executive that exercises the authority, while the nominal form does not need one and is not a predicate. Similar nouns are 旅游 (“travel or tourism”), 健身 (“exercise or fitness”), 纺织 (“weave or textile”), etc. For some nouns, some of their senses are nominalizations while others are not. For example, 教授 can be used as a noun (“professor”) or a verb (“to teach”). When it is used as a title, e.g., 王教授 (“Professor Wang”), clearly it is not a predicate of any kind. On the other hand, in 英语教授 (“English professor”), it is a predicate that takes “English” as its argument. In fact it is closely related to its verbal form, e.g., 他/她 教授/teach “教授” (“He teaches English”). For the purpose of this project, we are only interested in nouns that are nominalizations of its verbal forms.

Even when a noun is a true nominalized predicate, not all of its modifiers are legitimate arguments or adjuncts of this predicate. Some modifiers can only co-occur with the nominalized form and cannot co-occur with its corresponding verbal form. We are only interested in arguments and adjuncts that can co-occur with both the nominal and verbal forms of the predicate. When making this judgment our criteria are semantic rather than syntactic. For example, an adverb that is an adjunct to a verb is almost always realized as an adjective when it modifies the nominalized predicate, but it is still considered to be an adjunct to the nominalized predicate even though its syntactic category has changed. For example, 充分/thorough” is an adjunct (ARGM-MNR) of the nominalized predicate “调查/survey” because it can also be used as an adverbial modifier of its verbal counterpart. This is illustrated in (1). Also notice that ARG1 is a noun phrase when the predicate is a verb while it is a prepositional phrase when the predicate is nominalized.

(1) a. 联合国工业发展组织

Table 1: Argument structure of nominalized predicates
"The UN Industry Development Organization recently thoroughly surveyed the status of the industrial projects in the Tumenjiang region."

b. 联合国工发组织工业项目现状进行了一次工业项目状况调查
"The UN Industry Development Organization recently conducted a thorough survey regarding the status of the industrial projects in Tumenjiang Region."

In Chinese due to the lack of the morphological variations, the adjectival and adverbial modifiers often share the same form, just as the nominal and verbal forms of a predicate do. The most reliable difference is their syntactic distribution. Certain modifiers of nominalized predicates are typically noun-specific and should not be marked as an argument or adjunct of the nominalized predicates. With a few exceptions, determiner phrases (DP) and quantity phrases (QP) modifying the nominalized predicates are not annotated as arguments or adjuncts of the nominalized predicate because they are not possible arguments or adjuncts of their verbal form. This is illustrated in Example 2:

(2) a. 最近的一项调查显示 ... recent DE one CL survey show ...
   "A recent survey shows ..."

b. ...，产生一些消极影响...
   "...，some negative effects arose."

Exceptions are made for determiner phrases and quantity phrases that can also be modifiers of verbs, e.g., "五次"/times", "三/three 天/days", etc. These modifiers are typically expressions of duration and frequency.

When the nominalized predicate is the head of a relative clause, the relative clause as a whole is generally not an argument or adjunct of this nominalized predicate. However, its arguments or adjuncts may be found inside the relative clause as in (3).

(3) 报告认为，...，报告认为，亚洲金融危机
   "The report believes the effect that the Asian financial crisis created on the world economy is more serious than what is previously estimated."

3. Argument distribution

This section describes the typical syntactic realizations of the argument structure of nominalized predicates and the attachment ambiguity that arises in determining their dependency relations.

3.1. Arguments inside NP

Predicate is head of NP In many cases all arguments of a nominalized predicate can be located within the NP in which it is the head. Syntactically the arguments are realized as modifiers of the predicate. In Chinese, with rare exceptions these modifiers are to the left of the head. Depending on their semantic relation to the predicate, they should either be tagged as an argument (ARGn) or an adjunct (ARGM). The argument / adjunct distinction is drawn along the same lines as the arguments and adjuncts of the verbs: the argument are selected by the predicate and thus must fulfill the selectional restrictions of the predicate. The ARG0 of "合作/cooperation", for example, must be of multiple parties and must be animate entities that are capable of cooperating. The adjuncts (ARGM), on the other hand, can modify a wide range of predicates. Nominalization is generally accompanied by the adjective/adverb conversion: verbal predicates are modified by adverbs while nominalized verbs are modified by adjectives, even though these modifiers share the same semantic content. Therefore, we use the same functional tags to categorize the ARGMs of the verbal and nominal predicates, independent of their syntactic category.

(4) 这一地区成为海... this one region become Straits two side
   "This region became the best place for scientific and technological development and trade cooperation."

Predicate is modifier of NP head There are cases where the nominalized predicate is the modifier of another noun that is the head of the entire noun phrase. When the head noun of this phrase is also a nominalized predicate, as illustrated in (5), distinguishing the arguments of the modifier from those of the head amounts to a form of ambiguity resolution:

(5) 这是浙江省制定的...
   "This is the planning outline for the foreign economic and trade cooperation during the Ninth Five-Year Plan that Zhejiang Province has prepared."
Occasionally the head of the NP that a nominalized predicate modifies is an argument to this nominalized predicate. This happens when the predicate functions as a reduced relative clause, as illustrated in (6):

(6) [REL 调查] [ARG1 对象] 为 北京
    该城区十四至三十五岁的
    青年。

"The subjects of this survey are youths between fourteen and thirty-five years old living the Beijing Metropolitan area."

3.2. Predicate is subject

One characteristic of Chinese is the wide-spread use of topic constructions. When a nominalized predicate occupies the subject position, the topic is often an argument of this predicate. This is illustrated in (7).

(7) 近年 来，[ARG1 朝 韩 两
    国 之间 的 经营 往来 ]
    现在中日经济贸易合作
    [REL 发展] 迅速，
    [SUP 综合] 发展 稳定。

"In recently years, the economic and trade exchanges between China and South Korea developed rapidly."

3.3. Predicate occurs with a support verb

As is often the case, the nominalized predicate occurs with a support verb. Some support verbs have little or no semantic content and are generally there to fulfill a syntactic function. Other support verbs do add meaning to the nominalized predicate:

(8) a. 日本 还 应中国 有关 部门 之 邀
    日本也于在中国有关机关的邀请，对
    [ARG1 中国 长春 至珲春]
    由长春至珲春
    [ARGM-MNR] 两
    国之间的 经贸 往来 ]
    [REL 调查] [ARG1 日本 韩 综合] 迅速，
    [SUP 开发] [ARG0 德国] 形成 稳定。

"Japan also conducted a survey on the comprehensive development of the area along the railway from Changchun to Huichun, at the invitation of the relevant agencies of China."

b. [ARGM-LOC 在国际 事务 中 ]，
    [ARG1 欧盟 中国] [SUP 欧盟和中国]
    [ARG0 中国] [ARGM-MNR 很好] 的
    合作 [REL 合作]。

"China and foreign nations cooperated in many ways in industry, agriculture, trade, culture and education and other areas."
It is also worth noting that this particular support verb has its own semantic content.

(11) 进一步 [SUP 扩大] [ARGM-DIR 对] [through further expand] [toward] [outside] [REL 合作] [outside cooperation]...

"through further expanding foreign cooperation..."

It is possible that a verb is a support verb in some cases but not in others. A verb should not be marked as a support verb when it does not share an argument with the nominalized predicate, even though it may share an argument in other contexts.

3.4. Predicate is inside a prepositional phrase

When a nominalized predicate occurs in an NP that is the complement of a prepositional phrase, its argument can generally be found outside the prepositional phrase. This argument is often the (logical) subject of the matrix clause.

(12) 这份报告是 [ARGM-TMP 近 6 个月] [DE 学者] [through nearly 6 CL month] [的研究] [after propose] [DE].

"This report was presented by well-known researchers and scholars from Harvard University and University of California after six months of study."

Whether the argument of a nominalized predicate can be found outside the prepositional phrase also depends on the preposition. Some types of prepositions do not allow arguments of the nominalized predicate to be found outside the prepositional phrase:

(13) 海关工作要 [ARGM-MNR 全面] [DE] [ARG1 国家] [national treasury] [组成] [recently] [对] [ARGM-TMP 近] [ARG1 项目] [recently conducted a thorough survey regarding].

"The UN Industry Development Organization recently conducted a through survey regarding the status of the industrial projects in Tumenjiang Region."

b. 此间最新一项 [ARG1 投资] [ARGM-DIR 近] [ARGM-TMP 最近] [ARG1 环境] [here latest one CL] [investment].

A latest local survey on investment environment shows that investors of US, Japan, and EU, etc. widely believe that Shanghai’s comprehensive conditions for Shanghai has an obvious advantage in the comprehensive conditions of investment environment.”

Related to the event / result distinction is the phenomenon of incorporation, where the nominalized predicate absorbs one of its argument. When this happens the nominalized predicate takes on the meaning of both the predicate and the argument, and loses the event interpretation. In this sense it is similar to the result interpretation of a nominalized predicate. Unlike the result interpretation, however, incorporation does change the argument structure. When an argument is incorporated into the nominalized predicate, the nominalized predicate is labeled as both the REL and the label of the argument it incorporates. For any given predicate, incorporation may occur in some cases but not in others. (15a) illustrates the incorporation of an argument of “投资/invest” into the nominalized predicate while (15b) is an example where that no incorporation has taken place.

4. Event / Result distinction and argument incorporation

The argument structure of a nominalized predicate is largely independent of whether it denotes an event or a result in the sense that the number and type of arguments the predicate takes does not change regardless of whether the predicate has an event or result interpretation. 调查 has an event reading in (14a) indicating the act of the survey. In (14b), on the other hand, it has a result reading, indicating the result of the survey. However, it both cases it has two expected arguments, ARG0 the agent that conducts the survey and ARG1 the target of the survey, even though only ARG1 is realized in (14b). The event/result distinction is clearly an important one but since the main concern of this project is the argument structure, it is set aside for future work.

(14) a. [ARGM-MNR 联合国开发] [ARGM-TMP 近] [ARG1 组织] [investment environment] [here latest one CL].

b. 此间最新一项 [ARG1 投资] [ARGM-TMP 最近] [ARG1 环境] [here latest one CL].

"A latest local survey on investment environment shows that investors of US, Japan, and EU, etc. widely believe that Shanghai’s comprehensive conditions for Shanghai has an obvious advantage in the comprehensive conditions of investment environment.”

Related to the event / result distinction is the phenomenon of incorporation, where the nominalized predicate absorbs one of its argument. When this happens the nominalized predicate takes on the meaning of both the predicate and the argument, and loses the event interpretation. In this sense it is similar to the result interpretation of a nominalized predicate. Unlike the result interpretation, however, incorporation does change the argument structure. When an argument is incorporated into the nominalized predicate, the nominalized predicate is labeled as both the REL and the label of the argument it incorporates. For any given predicate, incorporation may occur in some cases but not in others. (15a) illustrates the incorporation of an argument of “投资/invest” into the nominalized predicate while (15b) is an example where that no incorporation has taken place.
The issues we have discussed so far are surprisingly similar to those described in the English Nombank guidelines (Meyers, 2005) where they discussed issues related to the annotation of nominalized predicates for the English Nom- bank. There are also notable differences between the two languages. The first has to do with the fact that Chinese verbs and their nominalizations share the same form. On purely morphological grounds, there seems to be little evidence to support the noun-verb distinction for the same Chinese word. As a result, the concept of nominalization is not one without controversy in Chinese linguistics, especially the Chinese language processing community. However, as we have demonstrated in Section 3., the arguments of verbs and their nominalizations are realized very differently. This lends support to the position that there are indeed differences between Chinese verbs and their nominalizations. For practical purposes, the lack of a formal distinction between the nominal and verbal forms is conductive to our annotation effort. We are able to use the same lexical guidelines (called frame files) used in the annotation of verbs for the annotation of their nominalizations. The second difference lies in the relation between the nombank annotation and its underlying syntactic annotation in the treebank. One difference between Chinese and English is that in Chinese, with rare exceptions, a prepositional phrase that (syntactically) modifies a verb phrase generally precedes the verb phrase. This is the case even if the verbs takes a noun phrase as its object and there is a semantic dependency between the prepositional phrase and this object. In English this semantic dependency can be represented by attaching the prepositional phrase at different levels: it can be attached at the same level as the verb if the prepositional phrase modifies the verb or it can be attached at the same level as the noun phrase if it modifies the noun phrase. The option of representing the semantic dependency in syntax and attaching the prepositional phrase to the noun phrase is not available to Chinese because while the noun phrase follows the verb the prepositional phrase precedes the verb. This semantic dependency is better addressed in semantic annotation. This is illustrated in Table 1, where ARG1 of the nominalized predicate 规划(“planning” is the prepositional phrase to the left of the support verb 进行(“conduct”).

6. Summary

In this paper we described the Chinese Nombank Project, a project is closely related to the Chinese Proposition Bank. We outlined the scope of the project and discussed the markability of the nominalized predicates and their arguments. We then categorized the distribution of the arguments of nominalized predicates. We also discussed the relevance of the event/result distinction to the annotation of nominalized predicates and the phenomenon of incorporation. Finally we discussed some cross-linguistic differences between English and Chinese. The first phase of the project has been completed. Over twenty thousand instances of nominalized predicates have been annotated from the 500 thousand word CTB. This is comparison with over eighty thousand verbal predicate instances.

7. Acknowledgement

I would like to thank Martha Palmer for her comments on an early draft of this paper and more importantly for her un-wavering support for this line of research. I would also like to thank Zixin Jiang for his suggestions and tireless annotation effort during the course of the project. This work is funded by the NSF ITR via grant 130-1303-4-541984-XXXX-2000-1070.

8. References

A. Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinski, B. Young, and R. Grishman. 2004. The Nombank Project: An Interim Report. In Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation, Boston, Massachusetts.

Adam Meyers. 2005. Annotation guidelines for nombank - noun argument structure for propbank.

Martha Palmer, Dan Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. Computational Linguistics, 31(1).

Nianwen Xue and Martha Palmer. 2003. Annotating the Propositions in the Penn Chinese Treebank. In The Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing, Sapporo, Japan.

Nianwen Xue, Fei Xia, Fu dong Chiu, and Martha Palmer. 2005. The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus. Natural Language Engineering, 11(2):207–238.