

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-

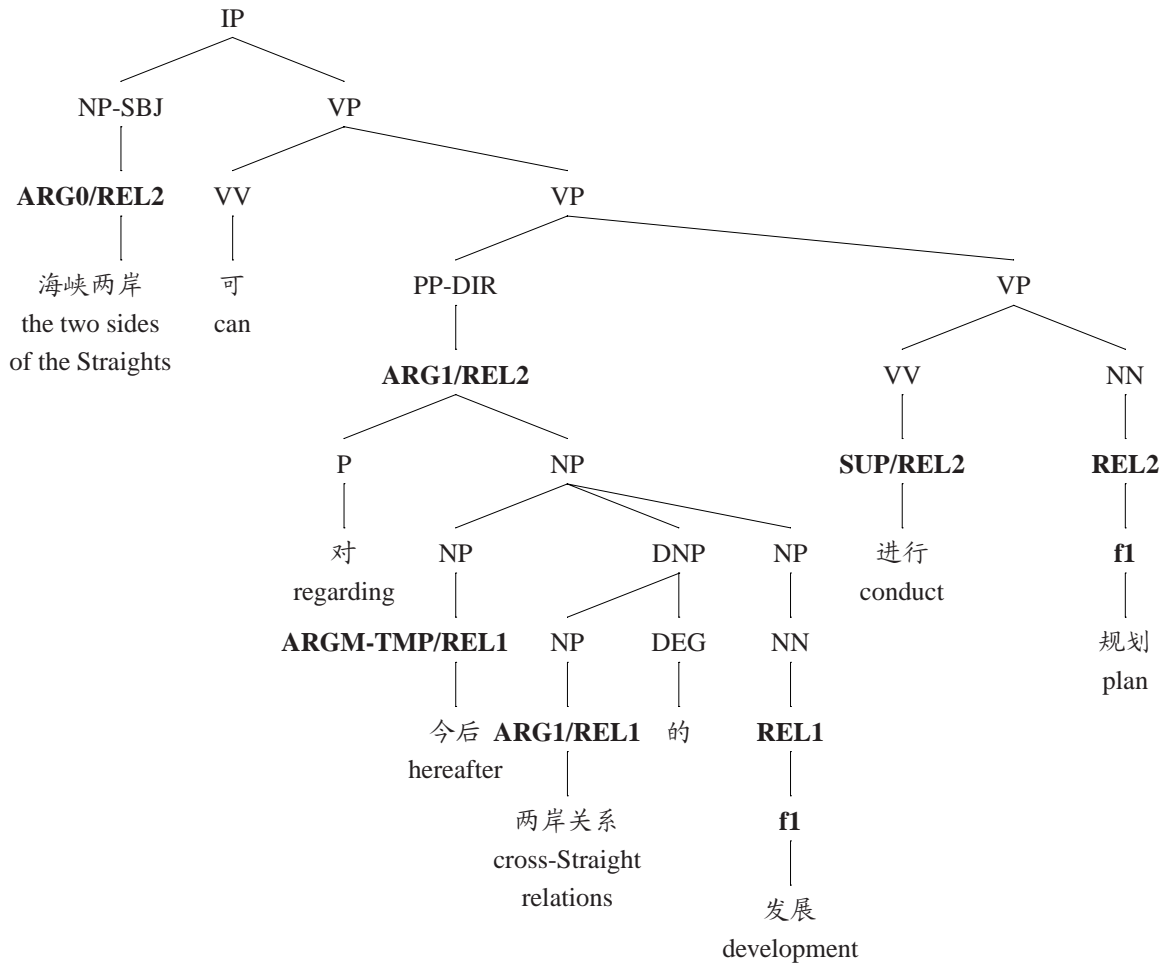
Strait relations"). "发展/development" also has a modifier labeled **ARGM-TMP**, 今后("hereafter"). The other nominalized predicate in this example, "规划/planning", has two possible arguments, **ARG0** the planner and **ARG1** the thing that's being planned. **ARG0** is "海峡/Straits 两/two 岸/sides" and **ARG1** is realized as a prepositional phrase "对/regarding 今后/hereafter 两岸关系/cross-Straits relations 的/DE 发展/development"

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 4. 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 form¹ with some verb and annotate the subset that

¹Unless otherwise noted, by morphological form we mean the way a word is written, not by how it is pronounced. This is admit-



The two sides of the Taiwan Straits can plan the development of the cross-Straight relations hereafter.

Table 1: Argument structure of nominalized predicates

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, they 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 英文教授 (“English professor”), it is a predicate that takes “英文/English” as its argument. In fact it is closely related to its verbal form, e.g., 他/he 教授/teach 英文/English (“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/thoroughly” 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. 联合国工发组织
 UN industry development organization
 近期 [ARGM-MNR 充分] [REL 调查
 recently thoroughly survey
] 了 [ARG1 图们江 地区 工业
 ASP Tumenjiang region industrial
 项目 现状].
 project status

tedly an unorthodox use of the concept.

"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
UN industry development organization
近期 [ARG1 对 图们江 地区
recently regarding Tumenjiang region
工业 项目 现状] 进行 了
industrial project status conduct ASP
[ARGM-MNR 充分][REL 调查]
thorough survey

"The UN Industry Development Organization recently conducted a through 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 二 项 [REL 调查] 显示 ...
recent DE one CL survey show ...
"A recent survey shows ..."
b. ... , 产生 一些 [ARGM-MNR 消极] [REL
... , arise some negative
影响]。
effect
"..., some negative effects arose."

Exceptions are made for determiner phrases and quantity phrases that can also be modifiers of verbs, e.g., "五/five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 , [ARG0 亚洲 金融 危机]
report believe , Asian financial crisis
[ARG1 给 世界 经济] [SUP 造成] 的 [REL
on world economy create DE
影响] 比 以前 估计 的 严重 。
effect than previously estimate DE serious

"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 selectiona 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 [ARG0 海峡 两岸]
this one region become Straits two side
[ARG1 科技
scientific technological ,
经贸] [REL 合作] 的 最佳
economic and trade cooperation DE best
地带 。
place .

"This region became the best place for scientific and technological, economic 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 [ARGM-TMP
this be Zhejiang Province prepare DE
"九五" 期间] [ARG1 对
"Nineth Five-Year Plan" duration toward
外 经贸] [REL 发展]
outside economic and trade development
[head 规划] 纲要 。
planning outline .

"This is the planning outline for the foreign economic and trade development during the Nineth Five-Year Plan that Zhejiang Province has prepared."

