

# Annotation Guidelines for the Chinese Proposition Bank

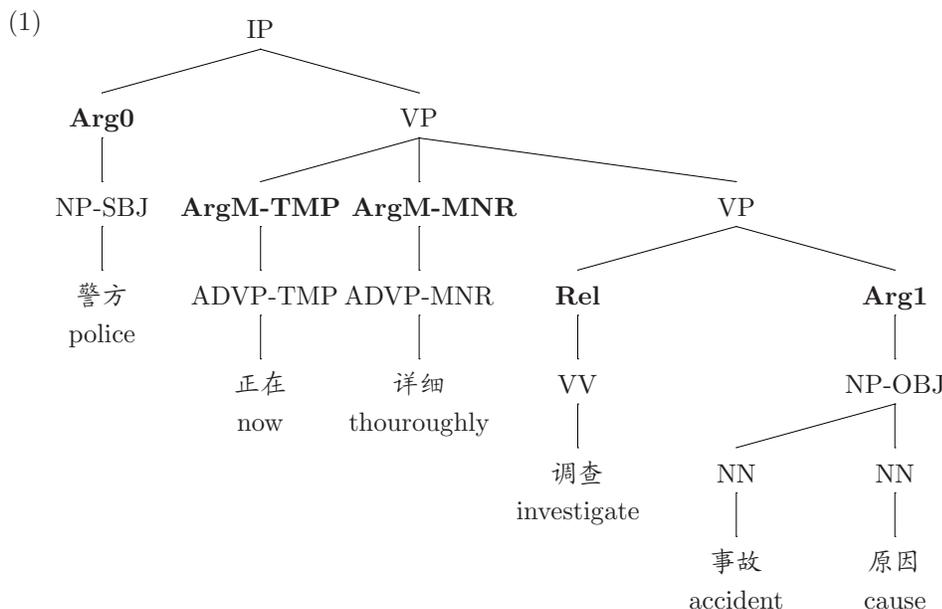
Nianwen Xue

February 19, 2007

## 1 Overview

### 1.1 Objective

The purpose of the Chinese PropBank (CPB) project is to add a layer of annotation to the hand-parsed sentences in the Chinese Treebank (CTB) (Xue et al., 2005). This layer of annotation assigns predicate-specific argument labels to the constituents in a parse tree. The arguments of each predicate in the sentence, which are limited to verbs and their nominalizations in the work we report here, receive an argument label in the form of *ArgN*, where *N* is an integer between 0 and 5. These numbered arguments represent *core* arguments that are defined in relation to the predicate, which is labeled as *Rel*. Each core argument plays a unique role with regard to the predicate and generally the total number of core arguments for each predicate does not exceed 6. The core arguments annotated for the verb 调查(“investigate”) in Example (1) are the NPs 警方(“the police”) and 事故(“accident”) 原因(“cause”), which are labeled as *Arg0* and *Arg1* respectively. The semantic role labels added to the parse tree are in bold.

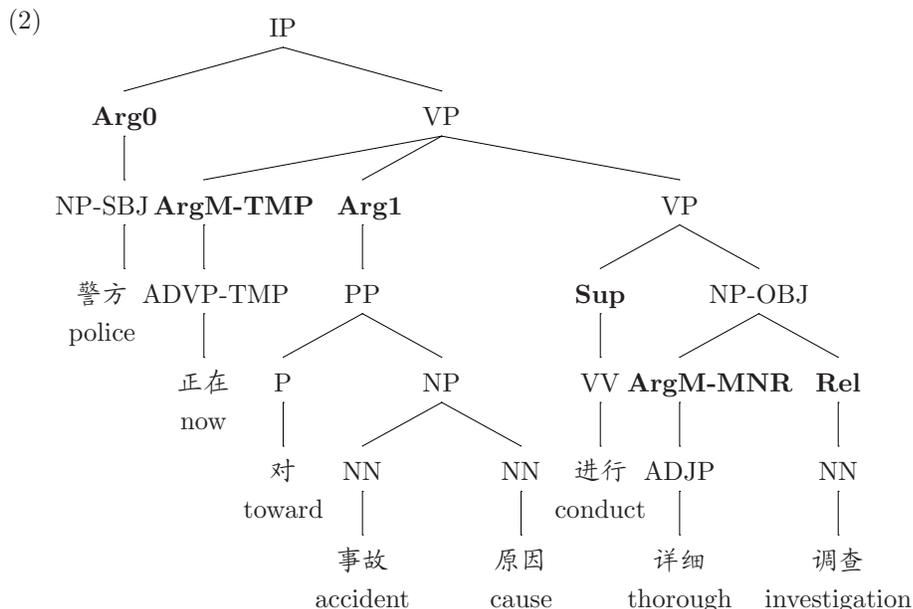


“The police are thoroughly investigating the cause of the accident.”

In addition to core arguments, we also annotate semantic adjuncts. Semantic adjuncts are not selected by the predicate and do not play unique roles with regard to a particular predicate. The flip side of the picture is that semantic adjuncts also do not tend to be selective as to which predicate they can occur with. They tend to represent general semantic notions such as temporality and location that are applicable to a heterogeneous set of predicates. It is not true, however, that temporal and locational elements are necessarily semantic adjuncts. For some predicates, temporal or locational elements may be required and are thus core arguments to the predicate. For example, 放, like its English translation “put”, requires a locational element

as one of its core arguments. Since semantic adjuncts are not required by a particular predicate, it makes little sense to define them in a predicate-specific manner. Therefore, the semantic role labels<sup>1</sup> assigned to semantic adjuncts are not predicate-specific and are represented in the form of *ArgM*, indicating adjunct, followed by a secondary tag representing the semantic classification of the adjunct. In (1), for example, the ADVP 正在 (“right now”), a temporal adjunct, is labeled *ArgM-TMP* and the ADVP 详细 (“thoroughly”) is marked *ArgM-MNR*. It should also be pointed out that the line drawn between arguments and adjuncts here is slightly different from what has been generally assumed in the theoretical linguistic literature, which is based on the obligatory/optional dichotomy. In a lot of cases, some constituents are labeled as core arguments even though they are clearly optional.

The predicate-specific argument labels of the Propbank annotation are designed to account for syntactic variations, the different ways in which the argument structure of a predicate can be realized. An argument is consistently labeled no matter how and where it is realized. In (2), a close paraphrase of (1), for example, the PP 对 (“toward”) 事故 (“accident”) 原因 (“cause”) still receives the semantic role label *Arg1* even though it undergoes a categorial change from NP to PP and is realized before the predicate 调查 (“investigation”) as a syntactic adjunct to the VP rather than as an object realized after the predicate. Notice also the categorial change of 调查 (“investigation”), a nominalized predicate in this sentence. In Chinese, nominalized predicates share the same form as their verb counterpart.



”The police are conducting a thorough investigation of the cause of the accident.”

Syntactic variations do not have to be accompanied by categorial changes of the predicate. In a book on English verbs, Levin (Levin, 1993) discussed extensively how each class of verbs can be realized in sets of subcategorization frames that are related through diathesis alternations. Similar alternations can also be found in Chinese. One common type of alternation is the “subject of intransitive / object of transitive” alternation that can be found in verbs that can be used both transitively and intransitively. When used intransitively the only realized argument is in the subject position. The same argument will occur in the object position in a transitive frame. The Propbank annotation scheme captures this by assigning the same semantic role label to this argument no matter where it occurs. 打开 (“open”) in (18) exemplifies a verb that can be used both transitively and intransitively. The noun phrase “中/China 美/the U.S. 交往/contact 的/DE 大门/door” occurs in the subject position in the intransitive frame<sup>2</sup> and in the object position when the verb is used transitively. In the semantic annotation discussed here, it is labeled *Arg1* in both frames, independent of its syntactic realization.

<sup>1</sup>Semantic role or argument labels refer to both core arguments and semantic adjuncts unless noted otherwise.

<sup>2</sup>In theoretic Chinese linguistics literature, this is often analyzed as a case where the subject is dropped and the object is fronted in a general topicalization process. We believe this is more appropriately analyzed as a case of syntactic alternation, tied to certain class of verbs.

- (3) a. [Arg1 中 美 交往 的 大门 ] [Rel打开] 了。  
 China the U.S. contact DE door open ASP .

”The door of contact between China and the U.S. opened.”

- b. [ArgM-TMP 7 0 年代初 ], [Arg0 中 美 两 国 领 导 人] [ArgM-ADV  
 70s beginning , China the U.S. two country leader  
 果断 ] [Rel 打开 ] 了 [Arg1 中 美 交 往 的 大 门 ]。  
 decisively open ASP China the U.S. contact DE door .

”In the beginning of the 1970s, the leaders of China and the U.S. decisively opened the door of contact.”

Some verbs can take on different sets of arguments that are realized in different sets of subcategorization frames. For example, the verb “存” expects two arguments when it means “exist”: the thing that exists and the location or domain in which it exists. When it means “deposit”, three arguments are expected: the entity that makes the deposit, the sum of money deposited and the financial institution which the deposit is made. When it means “preserve”, we expect three different arguments: the agent that does the preservation, the thing preserved and the instrument used in the preservation. Since each of these three senses can be realized in different subcategorization frames, in the propbank annotation convention, these senses are called *frameset*, meaning sets of subcategorization frames that realize a particular sense. The examples in (4) illustrate the three framesets of “存”.

(4) **Frameset 1:** “deposit”

Semantic roles:

*Arg0*: entity making deposit

*Arg1*: sum of money

*Arg2*: financial institution

- a. [ArgM-TMP 二十 年 前 ] [Arg0 每 人 ] [ArgM-ADV 平 均 ] [ArgM-ADV 才 ]  
 20 year ago each person average only  
 [Rel 存 ] [Arg1 二十 元 钱 ]。  
 deposit 20 yuan money .

“Twenty years ago, on average each person has only a deposit of 20 yuan.”

- b. [Arg1 大 批 资 金 ] [Rel 存 ] [Arg2 在 中 小 金 融 机 构 ]  
 large amount fund deposit in mid-sized small financial institution

“A large amount of funds are deposited in mid-sized or small institutions.”

**Frameset 2:** “exist”

Semantic roles:

*Arg0*: location

*Arg1*: thing that exists

- c. [ArgM-TMP 现 ] [Arg0 全 球 ] [ArgM-ADV 仅 ] [Rel 存 ] [Arg1  
 now the whole world only exist  
 一 千 多 只 大 熊 猫 ]。  
 over one thousand CL Giant Panda .

“There exist only a little more than one thousand giant pandas in the whole world.”

**Frameset 3:** “preserve”

semantic roles:

*Arg0*: preserver

*Arg1*: thing preserved

*Arg2*: instrument

- d. [Arg2 方志 ] 可以 [Rel 存 ] [Arg1 史 ], 资 治 , 教化 .  
 chronicle can preserve history , maintain order , civilize .

“Chronicles can be used to preserve history, maintain social order and teach civilized behavior.”

- e. [Arg0 我们] [Arg2 用 方志 ] [Rel 存 ] [Arg1 史 ], 资 治 , 教化 .  
 we with chronicle preserve history , maintain order , civilize .

“We can use preserve history, maintain social order and teach civilized behaviors with chronicles.”

## 1.2 Motivation

For some NLP applications, most notably information extraction and machine translation, it is desirable to extract the predicate argument (PA) structure of a verb, which specifies a relation, denoted by the verb, between its arguments. For example, in “John hit the window”, the predicate argument structure of the verb “hit” expresses a “hitting” relation between “John” and “the window”, which can be represented formally as *hit* (*John, the window*). While the hand-parsed sentences in the CTB provide a useful first step in getting to such predicate argument structures, they are insufficient due to the syntactic nature of their representations. There are three reasons why one cannot just read the predicate argument structure of a verb off a CTB parse tree.

### 1.2.1 Heterogeneity/Alternation of syntactic entities

First of all, it is often the case that one predicate argument structure can have multiple syntactic realizations in the form of *syntactic frames*. Certain arguments can be missing from some syntactic frames, and some arguments can appear in different syntactic positions in different syntactic frames. Which arguments can be missing and which arguments can alternate between different syntactic positions is often systematic, and driven by underlying semantic distinctions. This phenomenon is demonstrated extensively for English by Levin (1993) and similar alternations are also observed in Chinese, as illustrated in (5).

- (5) (IP (NP-SBJ (NN 新年/New Year)  
 (NN 招待会/reception))  
 (VP (NP-TMP (NT 今天/today))  
 (PP-LOC (P 在/at)  
 (NP-PN (NR 钓鱼台/Diaoyutai)  
 (NN 国宾馆/state guest house))))  
 (VP (VV 举行/hold))))

The New Year reception was held in Diaoyutai State Guest House today.

- (IP (NP-PN-SBJ (NR 唐家璇/Tang Jiaxuan))  
 (VP (NP-TMP (NT 今天/today))  
 (PP-LOC (P 在/at)  
 (NP-PN (NR 钓鱼台/Diaoyutai)  
 (NN 国宾馆/state guest house))))  
 (VP (VV 举行/hold)  
 (NP-OBJ (NN 新年/New Year)  
 (NN 招待会/reception))))

Tang Jiaxuan was holding the New Year Reception in Diaoyutai State Guest House today.

### 1.2.2 Syntax-semantics mismatches

The second problem in reading the predicate argument structure off a parsed tree is that the syntactic structure as specified in the CTB does not always transparently reflect the argument structure of a pred-

icate. To the extent that it does, certain syntactic entities have a unique mapping to an entity in the predicate-argument structure. This is not true for all cases. For example, the *syntactic adjuncts* in the CTB do not necessarily correspond to *semantic adjuncts*. A syntactic adjunct ”对/to 中华/China 人民/people 共和国/republic 中央/central 人民/people 政府/government 和/and 香港/Hong Kong 特别/special 行政区/administrative region” can be a semantic argument, in that it plays a role that is important to the verb ”负责/responsible”, as illustrated in (6). Another scenario where the syntax does not transparently reflect semantics is when certain verbs (as specified in the CTB) do not have their own predicate argument structure. Instead, they appear to introduce arguments of other verbs, as illustrated in (7), where 将/BA does not have a predicate argument structure of its own.

- (6) (IP (NP-SBJ (PN 本人/I))  
 (VP (PP-DIR (P 对/to)  
 (NP (NP (NP-PN (NR 中华/china)  
 (NN 人民/people)  
 (NN 共和国/republic))  
 (NP (NN 中央/central)  
 (NN 人民/people)  
 (NN 政府/government)))  
 (CC 和/and)  
 (NP-PN (NP-PN (NR 香港/Hong Kong))  
 (ADJP (JJ 特别/special))  
 (NP (NN 行政区/administrative region))))))  
 (VP (VV 负责/responsible))))

I am responsible to the Central People’s Government of the People’s Republic of China and Hong Kong Special Administrative Region.

- (7) (IP (NP-SBJ (NN 警察/police))  
 (VP (BA 将/BA)  
 (IP-OBJ (NP-SBJ-1 (QP (CD 三/three)  
 (CLP (M 名/CL)))  
 (ADJP (JJ 涉嫌/involve))  
 (NP (NN 犯罪/crime)  
 (NN 分子/element))))  
 (VP (VV 逮捕/arrest)  
 (NP-OBJ (-NONE- \*-1))))))

The police arrested three suspects.

### 1.2.3 Conflation of Predicate-Argument structures

Another reason why the PA structure is not transparently represented in a CTB parse tree is that the PA structures of multiple predicates are conflated in one syntactic frame in the case of some verb compounds. In this case, only the predicate argument structure of the verb compound as a whole is transparently represented and the PA structure of the component verbs have to be recovered. This is illustrated in (8).

- (8) (IP (NP-PN-SBJ (NN 公安部))  
 (VP (VCD (VV 开发)  
 (VV 研制))  
 (AS 了)  
 (NP-OBJ (PU “  
 (IP-APP (NP-SBJ (NP-PN (NR 港)  
 (NR 澳))  
 (NP (NN 居民)))  
 (VP (VV 来往))

(NP-OBJ (NN 内地))))  
(NP (NN 通行证))  
(PU " ))))

### 1.3 Scope of the Work

In this phase of the project we will annotate the PA structure of all verbs (including phrasal verbs and verb compounds, but excluding light verbs (e.g., *ba*, *bei*), raising verbs and auxiliaries). We also annotate the predicate argument structure of nominalized predicates. For verb compounds we annotate the PA structure of their component verbs, when they clearly have their own predicate argument structure. Initially we will focus on the six types of verb compounds explicitly marked in the Chinese Treebank: VSB, VRD, VCP, VPT, VNV, and VCD. We will not annotate the internal structure of verbs which could conceivably be recognized as verb compounds but have been segmented as one word in the CTB.

## 2 Annotation Procedure

The annotation of the predicate argument structure of a verb is carried out in two phases. In the first phase, the *frame file* for that verb is created. In the second phase, the predicate argument structure of all instances of that verb is annotated, using the frame file as a reference. This section describes how frames files are created and how the frame files are used in the actual annotation.

### 2.1 Creating frame files

The frame file for each predicate is created by examining all instances of the predicate in the Chinese Treebank. The sentences containing all instances of the predicate are extracted from the corpus and organized by the subcategorization frames using an extraction tool described in detailed in (Xue and Kulick, 2003). The frame file creator then examines these subcategorization frames and determines which set of subcategorization frames realize the same set of semantic roles. A frameset is then posited and the list of roles for that frameset are specified. This process is reiterated until all instances of the predicate are accounted for. A consequence of this frame file creation methodology is that it is possible that not all framesets of this predicate are exhausted. The worst case scenario is that these framesets need to be reorganized when new data needs to be annotated. From our experience, this worst case scenario rarely happens and new framesets can be added without affecting the existing framesets. In the next few sections we will examine the key elements in a frame file, namely, framesets, semantic roles and subcategorization frames, in greater detail.

#### 2.1.1 Framesets

Generally speaking each frameset can be viewed as a major sense of the predicate. The frameset is a necessary part of the predicate-argument annotation because different senses of a predicate may take different (core) arguments. The number of arguments for different senses of a predicate may be different, or even if the number of arguments is the same, they may be different types of arguments. For example, in (4), the "exist" sense has two core arguments while the "deposit" and "preserve" senses have three arguments. The sense distinctions can be made fairly consistently when the senses have different number of arguments, but the sense distinctions based on argument types can be murky and are known to be difficult to make. The key to consistent frameset determination is to identify reliable criteria based on which sense distinctions can be made. One such criterion is the diathesis alternations, the different ways in which the argument structure of a verb is realized. (Levin, 1993) examined extensively the diathesis alternations in English and used this information to classify English verbs based on the assumption that verbs that have similar diathesis alternation patterns are also semantically similar and they fall into semantically coherent classes. The flip side of this assumption is that if the verbs have different diathesis alternation patterns, they are also semantically different. Extending this assumption to verb senses, we expect that different senses of a verb may also exhibit different diathesis alternation patterns. One example that bear out this expectation is the verb 通过(9). All three senses "pass a bill or law", "pass through a tunnel", or "pass an exam or inspection" have two core arguments, but Frameset 1 has different alternation patterns from Frameset 2 and

3. Frameset 1 demonstrates the so-called "subject of intransitive / object of intransitive" alternation, i.e., 州际银行法("inter-state banking law") occurs in the object position in the transitive use of the verb and in the subject position in its intransitive use. Framesets 2 and 3, however, allow dropped object, but does not allow the "object of transitive/subject of intransitive alternation". The diathesis alternation cannot be used to distinguish all possible senses, however. For example, it does not distinguish Framesets 2 and 3. As a rule, the frameset distinctions for propbank annotation are coarse-grained and only made when necessary.

(9) Verb: 通过

**Frameset 1:** "pass (a bill, a law, etc.)"

美国 国会 最近 通过了 州际 银行法 。  
the U.S. Congress recently pass ASP inter-state banking law .

"The U.S. Congress recently passed the inter-state banking law."

州际 银行法 最近 通过了 。  
inter-state banking law recently pass ASP .

"The inter-state banking law recently passed."

**Frameset 2:** "pass through"

火车 正在通过 隧道 。  
train now pass through tunnel .

"The train is passing through the tunnel."

火车 正在通过 。  
train now pass through .

"The train is passing through."

**Frame 3:** "pass (an exam, etc.)"

他 没 通过 药检 。  
he not pass drug inspection .

"He did not pass the drug inspection."

他 没 通过 。  
he not pass .

"He didn't pass (He failed)."

One of main challenges in defining senses is the large number of dimensions along which senses can be partitioned. Even for the coarse-grained sense distinctions necessary for defining semantic roles there is no going around some language phenomena that influence sense determination. A few of them are discussed here.

**Idioms and Metaphors** One of the issues that we often have to deal with when differentiating senses for predicate argument annotation is idioms and metaphors. In the context of annotating the argument structure of predicates, idioms are cases where an argument of a predicate is always realized by the same word or phrases headed by the same word. Some examples are provided in (10). Idioms of this type generally require their own framesets when defining the semantic roles as they often have very different meanings from other uses of the same predicate.

(10) a. "establish (a precedent)"

它在向 联合国多边 机构 提供 援助方面 开 了一个很坏的  
it in for UN multilateral institution provide aid aspect establish ASP a CL very bad DE  
先例 。

precedent .  
"It established a bad antecedent for providing aid to UN multilateral institutions."

b. "make (a joke)"

中国 时报 驻 港 澳 主笔 谭志强 开 玩笑地 说  
China Times stationed in Hong Kong Macao chief correspondent Tan Zhiqiang make joke DE say  
。

."Chinese Times Hong Kong and Macao chief correspondent Tan Zhiqiang jokingly said."

Metaphors, on the other hand, generally do not call for a frameset of its own as they generally have a corresponding literal use that has a very similar argument structure. For example, the predicate-argument structure for the literal and metaphorical meanings of "drive" is the same (11a). The same is true for the literal and metaphorical meanings of "open" (11b).

(11) a. "drive"

一些 阔 太太开 着 宾士 而来 。

some rich lady drive ASP Benz come .  
"Some rich ladies came driving Benz."

白宫 关注 委 总统 对 民主 开 倒车 。

the White House concern Venezuela president toward democracy drive reverse car .  
"The White House is concerned that the Venezuela president is going backward on democracy."

b. "open"

警察 强行 开 门 ， 逮捕 嫌犯 。

police forcefully open door , arrest suspect .  
"The police forcefully opened the door and arrested the suspect."

政府 对 其 研发 大 开 方便 之 门 。

government toward its research and development widely open convenience DE door .  
"The government widely opened its door of convenience for its research and development efforts."

**light verbs** Light verbs are relevant to sense determination when a verb has both light verb and non-light verb uses. The argument structure for light verb and non-light verb uses are generally different. In a light verb construction, the real predicate is generally the nominalized predicate that the light verb supports. For example, in (12a) 作 is a light verb supporting the nominalized predicate 演讲("talk") while it has little meaning itself and does not have an argument structure of its own. 作 can also be used as a non-light verb, as in (12b), where it means "act as" and takes two semantic roles, the actor and the function the actor assumes.

- (12) a. 他经常被邀请到世界著名学府作讲演。  
 he often BEI invite go world famous academic institution give talk .  
 "He is often invited to the famous academic institutions of the world to give talks."
- b. 企业每逢收购合并，必然会请投资银行作顾问。  
 business every time buy merge , inevitably will invite investment bank act as consultant .  
 "Every time businesses purchase or merge with other businesses, they will inevitably invite investment banks to act as their consultants."

**Phrasal verbs** Chinese is a language in which compounds form a large portion of its vocabulary. While most verb compounds are treated as one-word segments in the Chinese Treebank that requires its own frame file, some verb compounds are multi-word expressions that are typically segmented as a sequence of two words. These multi-segment word compounds fall into three categories, based on the semantic relationship between the two component verbs. The two verbs can be coordinated and neither one is the head (e.g., 开发生产"develop and produce"), in which case the two verbs are each annotated with its own predicate-argument structure. The valency lexicon will thus have an entry for each verb, but not the entire compound. In the second type of compounds, the second verb is the head and the first verb is its modifier (e.g., 登记参选"register and become a candidate"). In this case only the second verb is annotated with its argument structure and a frame file is created for it. Again, the compound as a whole does not form a stable lexical unit and thus is not an entry in the valency lexicon. In the third type of verb compounds the head is the first verb and the second verb functions more as a particle than a full-fledged verb in the sense that it is semantically "bleached". It also falls into a closed class of verbs that can occur in this position and forms a more or less stable lexical unit with the first verb. These properties suggest that this third type of verb compounds is more like the phrasal verbs in English even though they are often described as verb resultative compounds in Chinese linguistics literature. Some examples are 建设成("construct-into"), 打扮成("dress-into") 看成是("consider-as"), 添加到("add-to"), 增加到("increase-to"), 提高到("improve-to"), 推迟至("postpone-to") 集中在("concentrate at"), 部署在("deploy at"), 凌驾于("perch at"),

In the Chinese Proposition Bank, these phrasal verbs are not treated as separate entries. In other words, they do not have their own frame files. Rather they are treated as a frameset in the frame file for the head verb. An example is given in (13). Generally speaking, the phrasal verbs have a different argument structure from that of the head verb, as the second verb generally introduces an argument.

- (13) a. Frameset: f1  
 ARG0: agent  
 ARG1: thing expanded  
 天津市 政府 准备 扩建 天津 滨海 国际 机场 。  
 Tianjin municipality government intend expand Tianjin Binhai International Airport .  
 " Tianjin municipality government intends to expand the Tianjin Binhai International Airport."
- b. Frameset: f2  
 ARG0: agent  
 ARG1: thing expanded  
 ARG2: thing expanded into  
 海拉尔 飞机场 [ 扩建 成为] 国际 航空 港 。  
 Hailaer Airport expand into international navigation port .  
 "Hailaer Airport has expanded into an international navigation port."





- b. [Arg1 这 名 考 生 ] 杭 州 大 学 没 有 录 取 。  
 this CL student Hangzhou University not accept .  
 ”This student Hangzhou University did not accept.”
- c. 杭 州 大 学 把 [Arg1 这 名 考 生 ] 录 取 了 。  
 Hangzhou University BA this CL student accept ASP .  
 ”Hangzhou University accepted this student.”
- d. [Arg1 这 名 考 生 ] 被 杭 州 大 学 录 取 了 。  
 this CL student BEI Hangzhou University accept ASP .  
 ”This student was accepted by Hangzhou University.”

**Syntactic alternations** One common type of alternation is the ”subject of intransitive / object of transitive” alternation that can be found in verbs that are used both transitively and intransitively. Verbs that demonstrate this alternation have two semantic roles, *Arg0* the agent and *Arg1* the patient or theme. When used intransitively only *Arg1* is realized and it occurs in the subject position. When used transitively, both *Arg0* and *Arg1* are realized, with the former in the subject position and the latter in the object position. The Propbank annotation scheme captures this syntactic variation by assigning the same semantic role label *Arg1* to the patient or theme no matter where it occurs. One of the verbs that demonstrate this alternation pattern is 打开(”open”) (18). The noun phrase 中(”China”) 美(”the U.S.”) 交往(”contact”) 的(DE) 大门(”door”) occurs in the subject position in the intransitive frame<sup>4</sup> and in the object position when the verb is used transitively. In the semantic annotation discussed here, it is labeled *Arg1* in both frames, independent of its syntactic position. Other verbs that demonstrate this alternation pattern include 改变(”change”), 改善(”improve”), 更改(”modify”), 关闭(”close”), 合并(”merge”), 化解(”resolve”), 缓和(”ease”), 解决(”solve”), 启动(”start”), 实施(”implement”), 实现(”realize”), 稳定(”stabilize”), 出版(”publish”), 发表(”publish”), 成立(”establish”), 批准(”approve”).

- (18) a. [Arg1 中 美 交 往 的 大 门 ] [Rel 打 开 ] 了 。  
 China the U.S. contact DE door open ASP .  
 ”The door of contact between China and the U.S. opened.”
- b. [ArgM-TMP 7 0 年 代 初 ], [Arg0 中 美 两 国 领 导 人 ] [ArgM-ADV  
 70s beginning , China the U.S. two country leader  
 果 断 ] [Rel 打 开 ] 了 [Arg1 中 美 交 往 的 大 门 ] 。  
 decisively open ASP China the U.S. contact DE door .  
 ”In the beginning of the 1970s, the leaders of China and the U.S. decisively opened the door of contact.”

放(”put, place”) exemplifies a class of verbs that take three arguments: an agent (*Arg0*), a patient or theme (*Arg1*), and a location (*Arg2*). Verbs of this class are generally variations of generic 放(”put, place”), and they differ in the manner of the putting action, the thing that is being placed or the shape and size of the location. Verbs of this class demonstrate a three-way alternation. In (19a), all three arguments are realized, with *Arg0* as the subject, *Arg1* as the object and *Arg2* as a postverbal adjunct. In (19b), *Arg0* is dropped and *Arg2* is realized as the subject while *Arg1* is still realized as the object. In (19c), *Arg0* is still missing, but *Arg1* is realized as the subject while *Arg2* is realized as a post-verbal adjunct. Verbs that demonstrate this alternation pattern include 挂(”hang”), 放置(”place”), 安放(”place”), 存放(”place and deposit”), 存(”deposit”), 盛(”fill”), 盛满(”fill”), 装(”load”), 装满(”load to full”), 装载(”load and carry”), 贴满(”stick to”), 摆放(”place”) and they all have these three arguments.

<sup>4</sup>In theoretic Chinese linguistics literature, this is sometimes analyzed as a case where the subject is dropped and the object is fronted in a general topicalization process. We believe this is more appropriately analyzed as a case of syntactic alternation, tied to certain class of verbs.

- (19) a. [Arg0 他] 把 [Arg1 一 台 电脑] [Rel 放] [Arg2 在 桌子上]。  
 he BA one CL computer put on desk LC .  
 "He put a computer on the desk"
- b. [Arg2 桌子上] [Rel 放] 了 [Arg1 一 台 电脑]。  
 desk LC put ASP one CL computer .  
 "A computer was put on the desk."
- c. [Arg1 一 台 电脑] [Rel 放] [Arg2 在 桌子上]。  
 one CL computer put on desk LC .  
 "A computer was put on the desk."

Verbs like 出现("appear") differ from verbs like 放("put, place") in that they do not allow an agent and only take two arguments, a theme *Arg0* and a location *Arg1*. There is again a three-way alternation. In (20a), *Arg0* the theme is the subject while *Arg1* the location is the object. In (20b), the positions of the two arguments are reversed: *Arg1* is in the subject position while *Arg0* is in the object position. In (20c) only *Arg0* the theme is realized and it is in the subject position. Verbs that demonstrate this alternation pattern include: 浮动("float"), 飘扬("flutter"), 漂浮("float"), 运行("operate"), 存在("exist"), 躲("hide"), 潜藏("hide"), 生活("live"), 坐("sit"), 活动("operate"), 旅居("reside"), 畅销("sell well"), 出现("appear"), 浮现("emerge"), etc.. Verbs that allow this alternation pattern have an existential sense. Non-existential senses of these verbs, if they exist, may not allow this alternation.

- (20) a. [Arg0 一 艘 船] [Rel 出现] [Arg1 在 地平线 上]。  
 one CL ship appear at horizon LC .  
 "A ship appeared on the horizon."
- b. [Arg1 地平线 上] [Rel 出现] [Arg0 一 艘 船]。  
 horizon LC appear one CL ship .  
 "A ship appeared on the horizon."
- c. [Arg0 一 艘 船] [Rel 出现] 了。  
 one CL ship appear ASP .  
 "A ship appeared."

Verbs like 撤离("leave, withdraw") expect two arguments and can also be used transitively and intransitively. When used transitively, *Arg0* the agent is in the subject position and *Arg1*, which can be a theme, location or event, is in the object position. When used intransitively, *Arg0* is still in the subject position but *Arg1* is dropped. Although dropped arguments are very common in Chinese (Huang, 1989), but only a subset of the Chinese verbs allow dropped object in a "neutral" context. Verbs that allow dropped object include 出席("attend"), 注册("register"), 撤离("leave, withdraw"), 进入("enter"), 采访("interview"), 参与("participate"), 学习("study"), 演出("perform"), 投资("invest"), 赔偿("compensate for"), 考虑("think about").

- (21) a. [Arg0 最后 一 批 俄罗斯 军队] [Rel 撤离] [Arg1 德国]。  
 last one group Russia army withdraw Germany .  
 "The last group of Russian army left Germany."
- b. [Arg0 最后 一 批 俄罗斯 军队] [Rel 撤离]。  
 last one group Russia army withdraw .  
 "The last group of Russian army left Germany."

Verbs like 鼓励(“encourage”) also allow dropped object, but they differ from the previous group of verbs in that they expect three arguments, an agent *Arg0*, an patient *Arg1* as well as *Arg2* which denotes the action taken by *Arg1*, encouraged or discouraged by *Arg0*. Verbs that allow this alternation pattern include 帮助(“help”), 欢迎(“welcome”), 鼓励(“encourage”), 允许(“allow”), 建议(“suggest”), 制止(“prevent forcefully”), 警告(“warn”), 限制(“restrict”). Verbs of this group only allow dropped argument, but do not allow their arguments to be realized in different syntactic positions.

- (22) a. [Arg0 政府 ] [Rel 鼓励 ] [Arg1 个人 ] [Arg2 投资 服务业 ] 。  
 government encourage individual invest service sector .

”The government encourages individuals to invest in service sector.”

- b. [Arg0 政府 ] [rel 鼓励 ] [Arg2 投资 服务业 ] 。  
 government encourage invest service sector .

”The government encourages investing in service sector.”

The next group of verbs, exemplified by the verb 译成(“translate”), expect four arguments, an agent *Arg0*, a patient or theme *Arg1*, a starting state, status or location *Arg2* and an end state, status or location *Arg3*. Verbs in this group generally indicate a change of status, state, or location involving *Arg1*, induced by *Arg0*. This group of verbs allow a four-way alternation as illustrated in (23). The list of verbs that allow this alternation pattern include 发展(“develop”), 撤回(“withdraw to”), 贬至(“devalue to”), 变成(“change into”), 变为(“change into”), 扩展到(“expand to”), 转变为(“transform into”), 转移(“move”), 搬到(“move to”), 晋升(“promote”).

- (23) a. [Arg0 他 ] 把 [Arg1 这 本 书 ] [Arg2 从 英文 ] [Rel 译成 ] [Arg3 中文 ] 。  
 he BA this CL book from English translate into Chinese .

”He translated this book from English into Chinese.”

- b. [Arg0 他 ] 把 [Arg1 这 本 书 ] [Rel 译成 ] [Arg3 中文 ] 。  
 he BA this CL book translate into Chinese .

”He translated this book into Chinese.”

- c. [Arg1 这 本 书 ] 已经 [Arg2 从 英文 ] [Rel 译成 ] [Arg3 中文 ] 。  
 this CL book already from English translate into Chinese .

”This book has already been translated from English to Chinese.”

- d. [Arg1 这 本 书 ] 已经 [Rel 译成 ] [Arg3 中文 ] 。  
 this CL book already translate into Chinese .

”This book has already been translated into Chinese.”

The syntactic variations described in this section are surprisingly similar to the diathesis alternations (Levin, 1993) described in her work on English verbs, although further examinations are needed to determine whether these are really the same alternations.

## 2.2 Annotating a verb

The annotation will be carried out in the following steps:

- Upon seeing the verb that needs to be annotated, the first step is to determine the subcategorization frame of this verb. This means distinguishing the semantic arguments from the semantic adjuncts.
- The second step is to decide which frameset this subcategorization frame belongs to. To do this, one needs to first determine which sense the verb has in this context if the verb has more than one senses. The relevant framesets for the verb is listed in frame file for this verb.

- The third step is to check which semantic roles the arguments of the verb maps onto and this information is listed in the frame file for this verb.
- The next step is to find out which predicate argument structure this syntactic frame maps onto. The mapping from the syntactic frame to the predicate argument is also specified in the frame file for this verb.
- The previous three steps only handle the arguments. The next step is to assign an appropriate functional tag to each of the adjuncts for this verb.
- There are two steps in annotating a verb compound. First the verb compound as a whole is annotated, following the steps specified above. Second the component verbs of the compounds are annotated when warranted. Whether the component verbs are annotated or not depends on the type of verb compounds.

### 3 Special Problems

This section discusses how the arguments of certain verbs are systematically dislocated from their canonical positions in some grammatical constructions.

#### 3.1 BA-construction

In the Chinese Treebank, *ba* is treated as a special verb that takes an IP as a complement. The use of *ba* demonstrates two main patterns in the CTB, depending on the verb of the complement IP. In the first case, the verb in the complement IP is a transitive or ditransitive verb or verb compound. The IP in which *ba* is the main verb generally has a non-*ba* alternative that means roughly the same thing. In this case, the predicate-argument structure for the two alternatives are the same. This means that we do not consider *ba* as having the predicate argument of its own and the subject of *ba* is considered to be an argument of the complement IP.

```
(24) (IP (NP-SBJ (NN 警察))
      (VP (BA 将)
          (IP-OBJ (NP-SBJ-1 (QP (CD 三)
                             (CLP (M 名)))
                       (ADJP (JJ 涉嫌))
                       (NP (NN 犯罪)
                          (NN 分子)))
                (VP (VV 逮捕)
                    (NP-OBJ (-NONE- *-1)))))))
```

```
(IP (NP-SBJ (NN 警察))
  (VP (VV 逮捕)
      (NP-OBJ-1 (QP (CD 三)
                  (CLP (M 名)))
              (ADJP (JJ 涉嫌))
              (NP (NN 犯罪)
                 (NN 分子))))))
```

```
REL: 逮捕
arg0: 警察
arg1: 三名涉嫌犯罪分子
```

```
(25) (IP (NP-PN-SBJ (NR 王码)
              (NN 电脑)
              (NN 公司))
```

(VP (BA 把)  
 (IP-OBJ (NP-SBJ (CP (WHNP-1 (-NONE- \*OP\*))  
 (CP (IP (NP-SBJ (-NONE- \*T\*-1))  
 (VP (ADVP (AD 不))  
 (VP (VV 加密))))  
 (DEC 的)))  
 (NP (NN 软盘)))  
 (VP (VV 奉献给)  
 (NP-OBJ (NN 用户))))))

(IP (NP-PN-SBJ (NR 王码)  
 (NN 电脑)  
 (NN 公司))  
 (VP (VV 奉献给) (NP-IO (NN 用户))  
 (NP-OBJ (CP (WHNP-1 (-NONE- \*OP\*))  
 (CP (IP (NP-SBJ (-NONE- \*T\*-1))  
 (VP (ADVP (AD 不))  
 (VP (VV 加密))))  
 (DEC 的)))  
 (NP (NN 软盘))))))

REL: 奉献给  
 arg0: 王码电脑公司  
 arg1: 用户  
 arg2: 不加密的软盘

In the second case, the main verb of the IP complement of *ba* is treated as a verb compound in the CTB. This is different from the first case in that there is another argument following the verb compound, which is treated as the object of the verb compound in the CTB. There are several observations for this use of *ba*: (a) the subject of *ba* can still be considered to be the subject of the complement IP, as in the first case, (b) the subject of the complement IP can be considered to be an object of the first verb of the verb compound rather than the object as a whole, (c) the object of the IP complement is introduced by the second verb of the verb compound and (d) the *ba* construction can sometimes have a grammatical non-*ba* alternative, though not always. On a closer examination one can find that the second verb (a) is generally semantically "bleached" and (b) belongs to a closed class. In this sense the second verb is more like a preposition rather than a verb. The verb compound as a whole behaves more like a phrasal verb, with the second verb (or a preposition) introducing an additional argument.

(26) (IP (NP-PN-SBJ (NR 苏联/USSR))  
 (VP (BA 把/BA)  
 (IP-OBJ (NP-PN-SBJ (NR 东德/East Germany))  
 (VP (VV 建成/construct-become)  
 (NP-OBJ (CP (WHNP-1 (-NONE- \*OP\*))  
 (CP (IP (NP-SBJ (-NONE- \*T\*-1))  
 (VP (VV 对付/counter)  
 (NP-OBJ (NN 西方/West))))  
 (DEC 的/DE)))  
 (NP (NN 桥头堡/bridgehead))))))  
 (PU . ))

An awkward equivalent:

(IP (NP-PN-SBJ (NR 苏联/USSR))  
 (VP (VV 建/construct)  
 (NP-PN-OBJ (NR 东德/East Germany)))

(VP/PP (VV 成/become)  
 (NP-OBJ (CP (WHNP-1 (-NONE- \*OP\*))  
 (CP (IP (NP-SBJ (-NONE- \*T\*-1))  
 (VP (VV 对付/counter)  
 (NP-OBJ (NN 西方/West))))  
 (DEC 的/DE))))  
 (NP (NN 桥头堡/bridgehead))))  
 (PU 。 ))

REL: 建成/(construct sth. so that it) becomes  
 Arg0: 苏联/USSR  
 Arg1: 东德/East Germany  
 Arg2-into: 对付西方的桥头堡/the bridgehead  
 with which to counter the West

### 3.1.1 BEI-construction

Like *BA*, we consider *BEI* to be a light verb that does not have a predicate argument structure of its own. Also like the *BA*, the existence of *BEI* causes the systematic dislocation of certain arguments of the verb in the complement VP, IP, or CP of *BEI*. In general, the verb in the complement VP/IP/CP has to be transitive, ditransitive, or otherwise allow an object which is fronted. The subject of the verb can be missing, in the case of short *BEI* (tagged SB). In this case, *BEI* takes a VP as its complement. Otherwise it is a long *BEI*, which takes a IP or CP.

**Long-BEI** A long *BEI* (POS-tagged LB) takes an IP or CP as its complement and the subject of the IP or CP is not missing. It takes a CP complement when a dislocated object of the verb in the complement CP is explicitly represented as co-indexed with an operator adjoined to the CP in the CTB. This operator is coreferential with the subject of *BEI*, although this is not explicitly represented. When annotating the predicate argument structure of the verb in the complement CP, the null object is replaced with the subject of the long *BEI*, with which it is coreferential via the null operator coindexed with it. Example (27) illustrates the annotation of the predicate argument structure of a transitive verb and Example (28) illustrates the annotation of the predicate argument structure of a ditransitive verb and an object control verb.

(27) (IP (NP-TMP (NT 今年))  
 (PU , )  
 (NP-SBJ (PN 他))  
 (VP (LB 被)  
 (CP-OBJ (WHNP-2 (-NONE- \*OP\*))  
 (IP (NP-PN-SBJ (NR 杭州)  
 (NN 大学))  
 (VP (VV 录取)  
 (NP-OBJ (-NONE- \*T\*-2))))))  
 (PU 。 ))

REL: 录取  
 arg0: 杭州大学  
 arg1: 他  
 argM-TMP: 今年

(28) (IP (NP-SBJ (DP-1 (DT 这些))  
 (NP (NN 债券))  
 (VP (LB 被)

```

(IP-OBJ (NP-PN-SBJ (NR 穆迪)
                  (NN 公司)
                  (CC 和)
                  (NN 标准)
                  (NR 普尔)
                  (NN 公司))
      (VP (VV 给予)
          (NP-IO (-NONE- *T*-1))
          (NP-OBJ (DNP (LCP (NP (NN 投资级))
                          (LC 以上))
                  (DEG 的))
                  (NP (NN 评级)))))))))

```

REL: 给予  
 arg0: 穆迪公司和标准普尔公司  
 arg1: 这些债券  
 arg2: 投资级以上的评级

```

(IP (NP-SBJ (QP (CD 五)
               (NP (NN 人)))
      (VP (LB 被)
          (CP-OBJ (WHNP-1 (-NONE- *OP*))
                  (IP (NP-PN-SBJ (NN 廉政)
                              (NN 公署))
                      (VP (VV 指控)
                          (NP-OBJ (-NONE- *T*-1))
                          (IP (NP-SBJ (-NONE- *PRO*))
                              (VP (VV 接受)
                                  (NP-OBJ (NN 贿赂))))))))))

```

REL: 指控 arg0: 廉政公署  
 arg1: 五人  
 arg2-PRD: 接受贿赂

Another use of long BEI is when a phrasal verb occurs in the complement IP of BEI. In this case the dislocation is not explicitly represented in the CTB. In general, the subject of the long BEI is the object of the head (first) verb of the phrasal verb / verb compound. The object of the phrasal verb is an extra argument introduced by the second verb (or preposition). Example (29) illustrates how the predicate argument structure of the phrasal verb is annotated.

```

(29) (IP (NP-SBJ (NN 植被)
      (VP (LB 被)
          (IP-OBJ (NP-SBJ (ADJP (JJ 大))
                      (NP (NN 车)))
                  (VP (VV 压成)
                      (AS 了)
                      (NP-OBJ (NN 路面))))))

```

REL: 压成  
 ASP: 了  
 arg0: 大车  
 arg1: 植被  
 arg2-INTO: 路面

### 3.1.2 Short-BEI

The use of the short-BEI is similar to the long BEI in many aspects. The most notable difference is that the subject of complement clause (annotated as a VP) is missing. The subject of the BEI is co-indexed with the null object of the verb of the complement VP when the verb is a transitive verb (Example 30), a ditransitive verb (31) or an object control verb (32). How the arguments are labeled is specified in the frame file for that verb <sup>5</sup>.

(30) (IP (NP-SBJ-3 (NP (NP-PN (NR 美国))  
(PP (P 对)  
(NP-PN (NR 华))))  
(NP (NN 贸易)))  
(NP (NN 逆差)))  
(VP (SB 被)  
(VP (ADVP (AD 严重))  
(VP (VV 夸大)  
(NP-OBJ (-NONE- \*-3))))))

REL: 夸大  
arg1: 美国对华贸易逆差  
argM: 严重

(31) (IP (NP-PN-SBJ-1 (NR 费孝通))  
(VP (SB 被)  
(VP (VV 授予)  
(NP-IO (-NONE- \*-1))  
(NP-PN-OBJ (NR 麦格赛赛奖))))))

REL: 授予  
arg1: 费孝通  
arg2: 麦格赛赛奖

(32) (IP (NP-SBJ-1 (NN 金融)  
(NN 部门))  
(VP (ADVP (AD 已))  
(VP (SB 被)  
(VP (VV 要求)  
(NP-OBJ-2 (-NONE- \*-1))  
(IP (NP-SBJ (-NONE- \*PRO\*-2))  
(VP (VV 开展)  
(NP-OBJ (NN 职工)  
(NN 个人)  
(NN 住房)  
(NN 抵押)  
(NN 贷款)  
(NN 业务))))))))))

REL: 要求 arg1: 金融部门  
Pred: 开展职工个人住房抵押贷款业务

---

<sup>5</sup>In general, the missing argument is arg0

Also like the long BEI, the short BEI can co-occur with phrasal verbs. The types of phrasal verbs / verb compounds are similar to those co-occurring with the BA and the long BEI. In this case, no explicit null category is posited in the CTB even though it is generally clear that the subject of the short BEI is the object of the head (first) verb of the phrasal verb. The second verb (preposition) introduces an extra argument which is annotated as the object in the CTB. The annotation of the predicate argument structure of this phrasal verb is similar to the long BEI construction except that the arg0 is missing. This is illustrated in (33). Words that co-occur with short-*bei* to take another argument: 列入, 列为, 确认为, 用於, 确定为, 誉为, 调往, 称为, 派往, 推举为, 聘请为, 推举为, 解职, 任命为, 推迟到, 介绍到, 选为, 认定为, 指定为, 命名为, 用于, 视为, 看作是, 审定为, 应用到, 称作, 认定, 纳入, 封为

(33) (IP (NP-SBJ (NN 住房))  
 (VP (ADVP (AD 却))  
 (VP (SB 被)  
 (VP (VV 当作)  
 (NP-OBJ (NN 职工)  
 (NN 福利))))))

REL: 当作  
 Arg1: 住房  
 Arg2: 职工福利

(IP (NP-PN-SBJ (NR 拉蒙·麦格赛赛奖))  
 (VP (SB 被)  
 (VP (VV 称为)  
 (NP-OBJ (DNP (NP-PN (NR 亚洲))  
 (DEG 的))  
 (NP-PN (NR 诺贝尔奖))))))  
 (PU 。))

REL: 称为  
 Arg1: 拉蒙·麦格赛赛奖  
 Arg2: 亚洲的诺贝尔奖

### 3.2 You-construction

For certain transitive verbs it is possible for the object to be fronted the subject position and the subject is introduced by the preposition 由. Such a construction generally has a non-由 alternative, as illustrated in (34).

(34) (IP (NP-SBJ-1 (DP (DT 这)  
 (CLP (M 次)))  
 (NP (NN 评选)))  
 (VP (PP-LGS (P 由)  
 (NP (NP (NP-PN (NR 中国))  
 (NP (NP (ADJP (JJ 对外))  
 (NP (NN 贸易)))  
 (NP (NN 经济)))  
 (NP (NN 合作部)))  
 (CC 和)  
 (NP-PN (NN 国家)  
 (NN 统计局))))))  
 (ADVP (AD 联合))  
 (VP (VV 组织)  
 (NP-OBJ (-NONE- \*-1))))  
 (PU 。))

```

(IP (NP-SBJ (NP (NP-PN (NR 中国))
                 (NP (NP (ADJP (JJ 对外))
                        (NP (NN 贸易)))
                    (NP (NN 经济)))
                (NP (NN 合作部)))
      (CC 和)
      (NP-PN (NN 国家)
              (NN 统计局)))
  (VP (ADVP (AD 联合))
      (VP (VV 组织)
          (NP-OBJ (DP (DT 这)
                    (CLP (M 次)))
                  (NP (NN 评选))))))
(PU 。 ))

```

REL: 组织

arg0: 中国对外贸易经济合作部和国家统计局联合组织

arg1: 这次评选

argM-MNR: 联合

## 4 Annotation of nominalized predicates

While the general annotation framework is the same for both verbs and their nominalizations, there are some issues that are specific to nominalized predicates. Certain nouns in Chinese are not true nominalizations even though there exists a verb that shares the same morphological form with them. 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. The ways in which arguments are realized are also 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.

### 4.1 What to annotate

To determine which nouns to annotate, we start by sifting through all nouns that shares the morphological form<sup>6</sup> with some verb and annotate the subset that 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

<sup>6</sup>Unless otherwise noted, by morphological form we mean the way a word is written, not by how it is pronounced. This is admittedly an unorthodox use of the concept.

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 (35). 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.

- (35) a. 联合国工发 组织 近期 [ArgM-MNR 充分 ][Rel 调查 ]了  
 UN industry development organization recently thoroughly survey ASP  
 [Arg1 图们江 地区 工业 项目 现状 ].  
 Tumenjiang region industrial project status

”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 近期 [Arg1 对 图们江 地区 工业  
 UN industry development organization recently regarding Tumenjiang region industrial  
 项目 现状 ]进行 了 [ArgM-MNR 充分 ][Rel 调查 ]  
 project status conduct ASP 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 36:

- (36) a. 最近 的 一 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 影响 ]。  
 ... , arise some negative 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 (37).

- (37) 报告 认为 , [Arg0 亚洲 金融 危机 ][Arg1 给 世界 经济 ][Sup 造成 ]的 [Rel  
 report believe , Asian financial crisis 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.”

## 4.2 Where to locate the arguments of nominalized predicates

### 4.2.1 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.

- (38) 这一地区成为 [Arg0 海峡两岸] [Arg1 科技、经贸]  
this one region become Straits two side scientific technological, economic and trade  
] [Rel 合作] 的最佳地带。  
cooperation DE best 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 (39), distinguishing the arguments of the modifier from those of the head amounts to a form of ambiguity resolution:

- (39) 这是浙江省 制定 的 [ArgM-TMP “九五” 期间] [Arg1  
this be Zhejiang Province prepare DE ”Nineth Five-Year Plan” duration  
对 外 经贸] [Rel 发展] [head 规划] 纲要。  
toward outside economic and trade development 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.”

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 (40):

- (40) [Rel 调查] [Arg1 对象] 为北京 城区 十四 至三十五 岁的 青年。  
survey subject be Beijing metropolitan area fourteen to thirty-five year DE youth.

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

### 4.2.2 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 (41).

- (41) 近年 来, [Arg1 中 韩 两国 之间 的 经贸 往来]  
recent years in, China South Korea two country between DE economic and trade exchange  
] [Rel 发展] 迅速。  
development rapid.

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

### 4.2.3 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:

- (42) a. 日本 还 应 中国 有 关 部 门 之 邀 ， 对 [Arg1 中国 长 春 至 珲 春 的 铁 路 沿 线 地 区 ] [ArgM-MNR 综 合 ] [Rel 开 发 ] [Sup 进 行 ] [Rel 调 查 ] 。  
 Japan also at China relevant agency DE invitation , toward China Changchun to Huichun DE railway along region comprehensive development conduct survey .

”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 ] , [Arg0 欧 盟 同 中 国 ] [Sup 进 行 ] 了 [ArgM-MNR 很 好 ] 的 [Rel 合 作 ] 。  
 at international affairs inside , European Union and China conduct LE very good DE cooperation .

”In international affairs, China and European Union also have had very good cooperation.”

Some support verbs are ”passive-oriented” and their presence is accompanied by a reversal of the arguments. One such example is 受(43):

- (43) 目 前 ， [Arg1 中 国 经 济 ] 更 易 [Sup 受 ] [Arg0 国 际 环 境 ] 的 影 响 。  
 presently , Chinese economy even easy get international environment DE influence .

” At present, Chinese economy is even more likely to be influenced by the international environment.”

By definition, a support verb is considered to be a support verb only if it at least shares an argument with the nominalized predicate. Whether the nominalized predicate also shares the adjunct modifiers of the support verb is a much more difficult question. This answer partly depends on the level of semantic content of the support verb and partly depends on the nature of the adjunct. In general, the less semantic content the support verb has, the more likely that the adjunct is licensed by the nominalized predicate. This is another area where much of the ambiguity resolution needs to be done. Our policy is to mark the adjunct as belonging to the nominalized predicate when it is clearly licensed by the nominalization, not the light verb. The (numbered) arguments are always marked. In (44a), the syntactic adjunct of the support verb ”加强/strengthen” is part of a split argument of the nominalized predicate ”合作/cooperation” and therefore must be marked as such. In (44b) the locative PP is clearly licensed by the nominalized predicate ”合作/cooperation” and therefore should be marked as such. This is also made possible by the fact that ”进行/conduct” has little semantic content.

- (44) a. [Arg0-CRD 东 盟 ] [Arg0-CRD 同 中 日 韩 ] [Sup 加 强 ] [Rel 合 作 ] 。  
 ASEAN with China Japan South Korea strengthen cooperation .

”ASEAN strengthens its cooperation with China, Japan and South Korea.”

- b. [Arg0 中 外 ] [ArgM-LOC 在 工 业 、 农 业 、 经 贸 、 文 教 等 方 面 ] [Sup 进 行 ] 了 { 多 种 } [Rel 合 作 ] 。  
 China and foreign countries in industry , agriculture , trade , culture and education etc. aspect conduct LE many kinds cooperation .

” China and foreign nations cooperated in many ways in industry, agriculture, trade, culture and education and other areas.”

In (45), it is clear that the underlined adverbial is licensed by the support verb 扩大, not by the nominalized predicate 合作. It is also worth noting that this particular support verb has its own semantic content.

- (45) 通过 进一步 [Sup 扩大 ] [ArgM-DIR 对 外 ] [Rel 合作 ] ...  
 through further expand toward 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 context.

#### 4.2.4 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.

- (46) 这份报告是 [Arg0 由美国哈佛大学 和 加州 大学的知名 专家  
 this CL report be by US Harvard University and California University DE well-known expert  
 学者 ] 经过 [ArgM-TMP 近 6 个月 ] 的 [Rel 研究 ] 后 提出 的 。  
 scholar through nearly six CL month DE research 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:

- (47) 海关工作 要 [P 为] [Arg1 国家 财政 和 经济 ] [Rel 建设 ] 作出 更  
 The Customs should to national treasury and economy development make even  
 大的 贡献 。  
 big DE contribution .  
 "The Customs should make an even bigger contribution to building the nation's treasury and economy."

We have not exhausted all possible context in which nominalized predicates can occur, but these are typical scenarios when nominalized predicates and their arguments can be found.

## References

- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language*, 67(3).
- Huang, James C.-T. 1989. Pro drop in Chinese, a generalized control approach. In Jaeggli O and K. Safir, editors, *The Null Subject Parameter*. D. Reidel Dordrecht.
- Levin, Beth. 1993. *English Verbs and Alternations: A Preliminary Investigation*. Chicago: The University of Chicago Press.
- Li, Charles and Sandra Thompson. 1976. Subject and Topic: A new typology of language. In Charles Li, editor, *Subject and Topic*. Academic Press.
- Xue, Nianwen and Seth Kulick. 2003. Automatic predicate-argument analysis of the penn chinese treebank. In *Proceedings of the Machine Translation Summit IX*, New Orleans, Louisiana, USA.
- Xue, Nianwen, Fei Xia, Fu dong Chiou, and Martha Palmer. 2005. The Penn Chinese TreeBank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, 11(2):207–238.