

Dealing with Nominalizations in Mandarin Chinese Using a Principles and Parameters Parser

Koong H.-C. Lin¹, Von-Wun Soo², and Sandiway Fong³

¹ Department of Business Administration, Ming Hsin Institute of Technology,
1 Hsin-Hsing Road, Hsin-Fong, Hsinchu, 304, Taiwan, R.O.C.

e-mail: koong@cs.nthu.edu.tw

² Department of Computer Science, National Tsing-Hua University,
Hsinchu, 30043, Taiwan, R.O.C.

e-mail: soo@cs.nthu.edu.tw

³ NEC Research Institute, 4 Independence Way, Princeton, NJ 08540

e-mail: sandiway@research.nj.nec.com

Abstract

This paper proposes how to extend Pappi, a principles and parameters parser which can currently parse ten languages based on the same core grammar, to handle nominalization constructions in Mandarin Chinese by deliberately revising the periphery file without affecting the other languages. There are four categories of nominalizations: relative clauses with a head noun as an adjunct, relative clauses with a head noun as a complement, nominalization with an abstract head noun, and nominalization with empty NP's and acts itself as an NP. We show analysis of such constructions from linguistic point of view based on the principles and parameters approach and its implementation. The experiments used a test corpus made up of sentences taken from judicial documents. A simple pattern matching mechanism was designed to extract sentences containing nominalizations. Experimental results show our work promising.

Keywords: Principles and Parameters approach, Government and Binding theory, theta theory, nominalizations, parser, Mandarin Chinese.

1. Introduction

In this paper, we study the difficult problem of parsing *Chinese nominal constructions* from Taipei Supreme court verdicts. This is part of a larger on-going effort to parse *real-world Chinese judicial documents*. Analysis of nominalizations is a hard issue in Chinese processing works. Most previous researches on designing Chinese processing systems only dealt with simple NPs; namely the study of nominalizations is still primitive. Lum & Pun [Lum & Pun, 1988] made a preliminary attempt to parse complex NPs including nominalization constructions. Their work was based on Case grammar and they implemented an *individual preprocessing module* which was rule-based for detecting complex NPs.

Chen [Chen, 1988] developed the first GB (Government and Binding theory) parser which can analyze some subsets of nominalizations (Although he didn't call them as "nominalizations"). Two types of nominalizations were handled: relative clauses with argument head nouns and nominalized constituents as appositives. The theta role assignment, constituent indexation, and semantic feature filling were handled; whilst the Case assignment was not discussed.

The work of Chang et al. [Chang & Lee, 1990], based on HPSG (Head-driven Phrase Structure Grammar), was a more complete work about nominalizations. Their analysis was lexicon-based, and based on Li and Thompson's classification of nominalization constructions [Li & Thompson, 1981]. They proposed a feature called "func" to specify declaratively the properties of the sentence constructions. Also a function-rule-firing mechanism was designed to process constructions involving functional words.

Dai's system ([Dai & Lee 1994], pg.67-76) was a revised version which improved Chang et al.'s work. He divided nominalization into two major types according to whether the nominalized structure had a head or not. Seven phrase structure rules were defined to cover nominalized sentences. The functional word "的 (de)" and slash features were the main information used. The feature values of nominalization without a head noun were obtained by using a structure sharing mechanism because there was no head daughter in the constituent structures. In the constituents of nominalization with a head

noun, the head features of the constituents were obtained according to Head Feature Convention.

We approach this nominalizations problem from the *Universal Grammar* perspective; in particular, as mentioned before, we adopt Chomsky's Parameters and Parameters Grammatical Theory (also known as GB) as outlined in [Chomsky, 1981; Chomsky, 1982; Fong, 1991; Huang, 1982; Lasnik & Uriagereka, 1988; Lin, 1995; Sells, 1985; Tang, 1992a; Tang, 1992b]. We will present a *linguistic analysis* of Chinese nominalization in this framework and discuss its *implementation* for parsing. We believe this work is significant for two reasons: (1) As far as we are aware of, this is the first complete account of parsing Chinese nominalization or complex NPs in this framework. As mentioned earlier, the existing works were based on Case grammar or HPSG. (2) Our parser is also the first principle-based parser to tackle *real-world* Chinese corpora. In fact, aside from PRINCIPAR [Lin, 1994] for English, this work represents one of the few efforts to build broad-coverage principle-based parsers. (3) The design of our parser preserves the modularity merit of the principles and parameters approach. PRINCIPAR mixes these principles into a message-passing framework, and thus forfeits its modularity nature. (4) In contrary to the other principle-based parsers, such as UNITRAN, the periphery file which describes the language-specific phenomena is implemented in our work.

We began by extending the PAPPI engine, an existing principle-based parser with a core set of principles parameterized for nine languages, to handle Chinese. This involves, *inter alia*, setting the PAPPI language parameters for Chinese; for example, to get the word order right. In particular, note that, unlike English, Chinese relative clauses must precede the head noun. An important advantage of this framework is that the universal nature of the *core grammar* allows us to leverage or re-use those principles already constructed for other languages to handle Chinese. In this paper, we focus on adding language-particular rules and generalizing the core principles to handle the special problems posed by the Chinese nominal constructions found in our judicial documents.

The phenomena of Chinese nominalization will be introduced in the next section. In section 3, we will briefly describe the principles and parameters approach; and propose our linguistic resolution for nominalizations based on this theory. Then the

implementation of the linguistic analysis is shown in section 4. Finally experimental results which attest to our promising work and conclusions are given in section 5.

2. Interactions with Chinese Nominalizations

In many languages, there are different processes which can make verbs, VPs, sentences, or portions of a sentence to be used as an NP (noun phrase). Such a process is called *nominalization*. In Chinese, a verb, a VP, a sentence, or a portion of a sentence is nominalized by adjoining the marker "*de*" ("的" or "之") behind it. Consider the following examples:

- S1.** 種 水果 的 農人 (grow fruit de farmer)
(the farmer who grows fruits)
- S2.** 我 寫 信 的 筆 (I write letter de pen)
(the pen I use to write letters)
- S3.** [[我們 合作] 的 問題] 很簡單 (we cooperation de problem very simple)
(The problem about our cooperation is very simple.)
- S4.** 這種 植物 可以 當作 [吃 的 e] (this kind plant can be taken as eat de e)
(This kind of plants can be taken as food.)

In **S1**, the CP "種 水果" becomes to function as an NP after the marker "de" is attached behind it. In addition, the CP also serves as a modifier for the adjacent noun "農人". Thus, the structure of a nominalized phrase is "[CP *de* NP]". **S2** is similar to **S1** in the sense that both serve as relative clauses. The major difference is that, the head noun "農人" in **S1** is a complement of the verb "種", while "筆" in **S2** is an adjunct of "寫" (Specifically, it's an instrument of "寫"). Again **S3** seems similar to **S1** and **S2** but in fact is quite different; the nominalization [[我們 合作] 的 問題] in **S3** is not a relative clause since "問題" is neither complement nor adjunct of "合作". In fact, the CP is an appositive clause of the head noun. The [吃 的] in **S4** is also a nominalized clause but the head noun is

empty; it means something which can be eaten. Sentences *S1-S4* exhibits the difficulties of Chinese nominalizations. Summarizing the above sentences and according to [Li & Thompson, 1981] and [Dai, 1994], the nominalization can be classified into four categories, as listed in table 2.1.

Table 2.1: Categories of Nominalizations

| <i>Type</i> | <i>Function of the nominalization</i> |
|---------------|---|
| <i>Type-1</i> | Relative clauses with head noun as a complement |
| <i>Type-2</i> | Relative clauses with head noun as an adjunct |
| <i>Type-3</i> | Nominalization with an abstract head noun |
| <i>Type-4</i> | Nominalization with empty NP's and acts itself as an NP |

3. Principle-based Approach to Parse Chinese Nominalizations

3.1. Basic Assumptions

The basic assumption of the principles and parameters approach is to assume the existence of a *universal grammar* that can account for the common characteristics of all languages. That is, any human language can be constructed by a *core grammar* and a *periphery grammar*. The core grammar consists of the universal principles of all languages, while the periphery grammar consists of the specialized rules of individual languages.

In the principles and parameters approach, the core grammar consists of a number of interesting principles that deal with different aspects of sentence interpretation. For example, the θ -theory module is concerned with how θ -roles such as agent and patient are properly assigned to phrases in a sentence. Similarly, Binding Theory deals with the distribution of anaphors and pronouns. Move- α is the transformation principle describing how to move a constituent to somewhere else. The locality of the domain

where syntactic phenomena occur is governed by Government Theory. Control Theory explains the distribution and relationships between PRO (empty subjects in infinitival clauses) and its controller in the control construction and Case Theory tells us about the Case assignment. Finally, phrase structure is described by X-bar Theory and Projection Principle.

The modules of grammar are parameterized with respect to individual languages. These language *parameters* encode systematic differences between the various languages. For a more in depth description of the principle and parameter theory, see [Chomsky, 1981; Fong, 1991; Huang, 1982; Lasnik & Uriagereka, 1988; Lin, 1995; Sells, 1985; Tang, 1992b].

3.2. Type-1: Relative Clauses with a Head Noun as a Complement

The constructions with a head noun modified by the nominalization referring to some unspecified participant in the CP are called relative clause constructions. For example, in [Ex. 3.2], "他們 種 的" modifies "水果", and "水果" refers to the object of "種", the verb of the CP "他們 種 *e*".

[Ex. 3.2] [NP [CP 他們 種 *e*] 的 水果]
(they grow de fruit)
(the fruits they grows)

Along the lines of the stipulation made in ([Huang, 1982],pg.485), the relativization process involves the movement of an *abstract operator*, called *OP*. Observing figure 1: In S1, the empty OP is in the complement position of the predicate "種 (grow)". Whilst in S2, OP moves from inside IP to the specifier position of CP, leaving a trace *t_i*, so that the scope of the operator OP is the IP. Thus, in the semantic interpretation, OP is a lambda predicate "(Lambda)" which takes the entire IP to describe the head noun " 水果 (fruit)".

S1: [CP [IP 他們(they) 種(grow) *OP*]] 的(de) 水果(fruit)
S2: [CP *OP_i* [IP 他們(they) 種(grow) *t_i*]] 的(de) 水果(fruit)

Figure 1: Movement of OP

Hence, our resolution for type-1 nominalization mainly focuses on: (1) allowing OP in the relative clause context. (2) constructing appropriate phrasal structures for relative clauses, as depicted in figure 2. Details for implementation will be described in section 4.1.

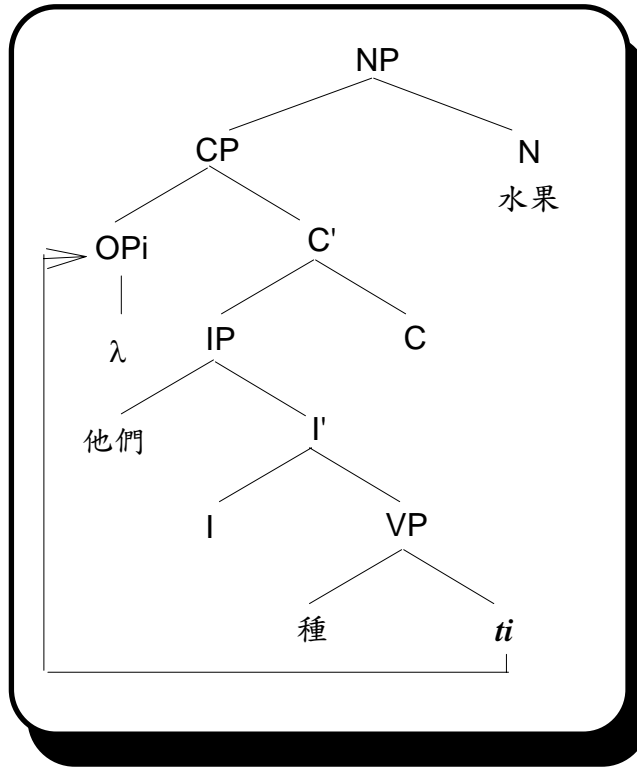


Figure 2: Phrase structures of relative clauses

3.3. Type-2: Relative Clauses with a Head Noun as an Adjunct

This kind of nominalization is very similar to what is mentioned in section 3.2. The only difference is that the head noun in section 3.2 is a complement of the verb in the CP; while the head noun here is an adjunct which often represents the concepts of *instrument*, *location*, *time*, *reason*, or *method*. For example, in [Ex. 3.3], "原因" refers to the *reason* of the action "I come here".

[Ex. 3.3] 我 來 這兒 的 原因

(I come here de reason)

(the reason why I come here)

For simplicity¹, we can handle it by stipulating that verbs that allow for an external instrument role can license the head NP to which the clause is adjoined to. Thus, our current resolution is registering the adjunct information in the lexical entries of verbs, and treat this kind of nominalizations as common relative clauses.

¹The only way we can make the instrument nominalization a relative clause construction is to have an empty operator. But if we have an empty operator, then it needs to have an empty category (ec) within its scope to license as a bound variable. If we have an ec which is not an obligatory argument, then it needs to be a VP adjunct. If it is a VP adjunct then it needs to be inside a PP. In which case, we need an empty P and so on. It gets complicated.

3.4. Type-3: Nominalization with an Abstract Head Noun

This type of nominalization is more complicated; since the head noun is neither the complement nor the adjunct of the main verb. Instead, the head noun is some kinds of abstract noun, such as "判決" (judgment), "事實" (fact), "事件" (event), etc., which abstracts the entire CP. That is, the head noun doesn't refer to any specified or unspecified entity; namely the nominalization is not a relative clause. And thus, the difficulty is how to find the relationship between the CP and head noun.

The classification of nominalization based on [Li & Thompson, 1981] and [Dai & Lee, 1994] needs to revise. We find that this category should be divided into two subcategories:

Type-3a: The nominalized constituent as a complement of the head noun

Type-3b: The nominalized constituent as an appositive clause of the head noun

In the following, we will analyze these two subcategories respectively.

Type-3a: The nominalized constituent as a complement of the head noun

In this category, the head noun belongs to some kind of words with both noun and verb syntactic categories, like "判決" (judgment), "宣告" (claim), and so on. Such words have the same semantics regardless of being treated as a noun or a verb. The nominalizing process is transforming the original [*Hd* CP] structure (Hd: the head noun with a verb category) into [CP de *Hd*]. For example, the original form of [Ex 3a-2] is [Ex 3a-1]. Therefore, we say that in this category, the CP is a complement of the head noun.

[Ex. 3a-1] [CP *PRO* 判決 [CP 被告 違 法]]

(judge the defendant violate law)
(Judging that the defendant violated the law)

[Ex. 3a-2] [NP [CP 被告 違 法] 的 判決]
(the defendant violate law de judgment)
(The judgment that the defendant violated the law)

In order for such constructions to be accepted, we think CPs should acquire Case. We find Tang also made a similar argument; he proposed that the requirement that NPs must have Case should be extended to the CPs for Chinese ([Lin & Soo, 1995], pg.24) ([Tang, 1992a], pg.336) since "的" (de) seems to be necessary for Case reasons. This argument is consistent with our analysis. Thus, we assert the following conjecture:

[Conjecture 3.4.1]

CP needs Case.

However, the highest CP in a sentence doesn't need Case. Therefore, conjecture 4.1 is revised as:

[Conjecture 3.4.2]

CP, *except matrix CP*², needs Case.

It's tempting to note that our conjecture dovetails Tsai's argument [Tsai, 1995]: "*CP arguments, as well as NP arguments, need Case to receive thematic roles*". Visibility condition posits that "theta roles may be assigned to A-positions³ which are associated with PRO⁴ or Case." [Aoun, 1979] [Chomsky, 1981]. In other words, argument NPs are visible for theta-marking only when they are Case-marked [Tsai, 1995].

²The matrix CP is the highest CP in a sentence.

³A-positions (argument positions) are potential theta positions, positions to which a theta role can be assigned [Haegeman, 1994].

Tsai has proved that since CPs and NPs are observed to have the same distribution in Chinese, it is likely CPs also need Case when construed as an argument. Tsai's opinion is consistent with conjecture 3.4.2.

Herein we take "被告 違法 之 判決" (the defendant violate law de judgment) (The judgment that the defendant violates the law.) as an example.

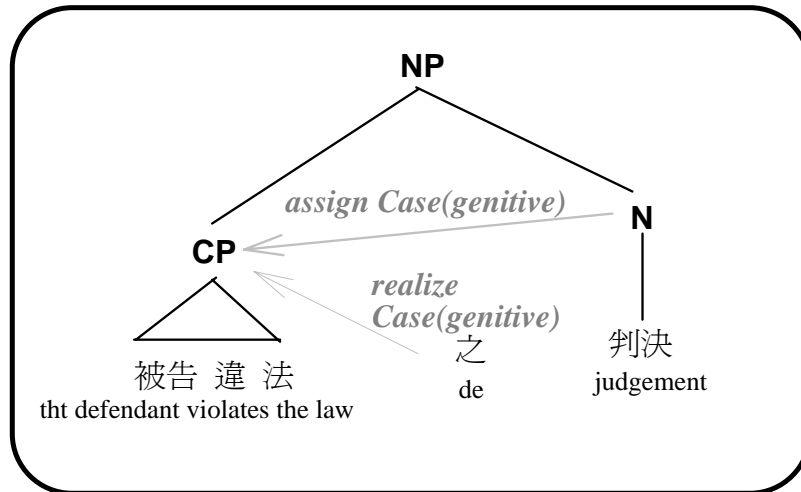


Figure 3: Assigning Case to CP

"被告 (the defendant)" (N), "違 (violate)" (V), and "法 (law)" (N) construct a CP. Since this is not a matrix CP, it needs Case. Thus, the noun "判決 (judgment)" following "之 (de)" assigns it a genitive Case. Note that the Case assignment is realized by the marker "之 (de)". After this process, the CP and the N combine together to form an NP. We can see that under this kind of nominalization, CP is an appositive clause which modifies its subsequent N.

Type-3b: The nominalized constituent as an appositive clause of the head noun

⁴ Infinitival clauses without overt subjects have a non-overt subject represented as PRO [Haegeman, 1994].

In this category, the CP is an appositive clause of the head noun which is always abstract, such as "事實" (fact), "問題" (problem), 新聞" (news), and so on. Note that our resolution for type-3a is not suitable for this category. Since the head noun assigns a Case to the CP, it should assign a theta role to the CP. However, if the CP is not an appositive clause of an N, the CP will not be N's argument (complement) and thus cannot receive any theta role.

There are two possible resolutions for type-3b:

- (1) CP is an adjunct of an N.
- (2) explain such construction in "DP" (Determiner Phrase), as in figure 4.

In such structure, "de" Case marks the CP; it assigns structural Case to the CP and doesn't need the CP to be an argument.

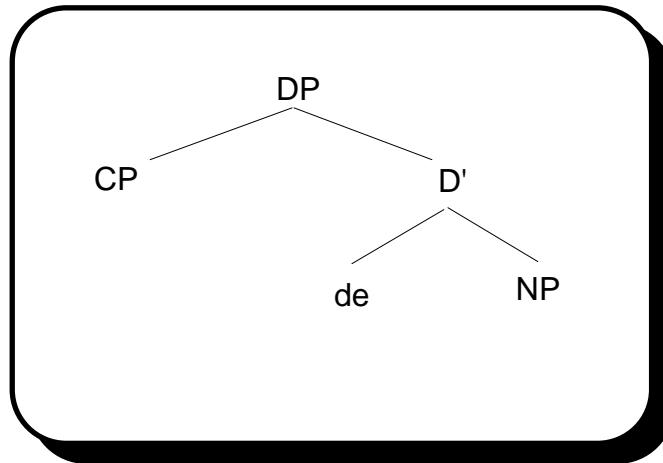


Figure 4: DP: a determiner phrase

However, just like QP, currently we haven't define the DP in Pappi yet. Thus, the former one is used as our current resolution.

3.5. Type-4: Nominalization with Empty NPs and Acts Itself as an NP

In such nominalization, the constituent following "de" is an empty NP. Namely, the head noun refers to the omitted entity of a CP is also missing. Therefore, we can describe such constructions as [CP de e], where "e" means "empty". For example, the nominalized NP in [Ex. 3.5] refers to the object of "like", i.e., the thing I like.

[Ex. 3.5] 你 沒 有 [NP [CP 我 喜歡 e] 的 e]

(you not have I like de e)

(You don't have what I like.)

This kind of nominalization can be handled by similar treatment as relative clauses; with an additional assumption:

[Conjecture 3.5]

An empty NP can assign Case to a relativized CP.

In addition, the reference of the omitted constituents should be determined. Summarizing from [Li & Thompson, 1981] about the nominalization [CP de e], in a CP, according to the theta grid registered in the lexical entry of V, there are three possibilities for missing NPs:

- (1) [*e* V *e*]: The nominalized NP refers to the missing *object*. (It is because Chinese is a *pro-drop* language. That is, the subject is often omitted.)
- (2) [*e* V *object*]: The nominalized NP refers to a *subject*.
- (3) [*subject* V *e*]: The nominalized NP refers to an *object*.

Case (2) and (3) are intuitive; and (1) means that objects has higher priority than subjects to be the referred targets. Besides, what about ditransitive verbs? Li and Thompson proposed the following argument [Li & Thompson, 1981]:

The nominalized NP doesn't refer to the indirect object.

That is, the indirect object will not move outside the CP and thus doesn't need to be considered. Nevertheless, we find counter examples:

- E1: [NP [CP 我 給 王小姐 *e*] 的 書]
 I give Miss Wang de book
 the book I gave Miss Wang
- E2: [NP [CP 我 給 王小姐 *e*] 的 *e*]
 I give Miss Wang
 something I gave Miss Wang

In E1 and E2, the indirect object "書 (book)" moves outside the CP, and thus violate the line stipulated by Li and Thompson. Moreover, it's still ambiguous when both objects are omitted:

- E3: [NP [CP 我 給 *e* 書] 的 *e*]
 I give book de
 somebody to whom I gave books
- E4: [NP [CP 我 給 *e e*] 的 *e*]
 I give de

E4 is indeed ambiguous; there are two possible interpretations: the nominalized construction may refer to either "something I gave Miss Wang" or "someone to whom I gave books". Therefore, more exploration is needed.

4. Chinese Nominalizations and Computational Implementation

Our parser, called Pappi (Principle and Parameter Parser), is designed based on the Principles and Parameters Grammatical Theory (also known as GB, Government and Binding Theory). It can currently handle ten languages: *English, Japanese, French, German, Dutch, Spanish, Korean, Bangla, Hindi, and Chinese*. Note that in Pappi, all of them are handled using the same core grammar; e.g., the same principle set and the same X-bar phrasal structures [Fong, 1991]. Hence, whenever dealing with a new phenomenon in one language, we must be very deliberate not to affect the other languages. The idiosyncrasy of individual language is reflected via setting its *parameter, periphery, and lexicon* files. Our focus here is to parse Mandarin Chinese sentences. Parameters for Chinese have been carefully set; for example, only C and N are head final, the Case adjacency, pro-drop, parameters are hold, the preposition stranding, wh-in-syntax parameters are not hold, with IP and NP being the subadjacency bounding nodes, and so on. Our goal is extending the periphery file for Chinese to handle nominalization constructions. Figure 5 shows a typical example of Chinese phrase structure as produced by the parser. In the subsequent sections, we will describe how to *computationalize* our linguistic analysis mentioned in section 3 based under the framework Pappi. Note that before nominal constructions are considered, the following theories exist in Pappi and need revisions later:

- (1) [Case Filter]: NPs with phonetic contents require Case [Chomsky, 1981].
- (2) The head noun in a relative clause must be overt.

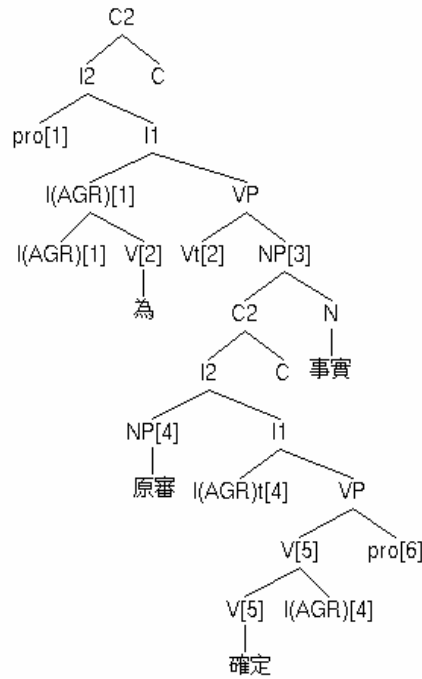


Figure 5: a typical example of Chinese phrase structure

4.1. Computational Implementation to Deal with Type-1 Nominalization

As analyzed in section 3.2, the major mission is to allow an empty operator (OP) in the relative clauses context. As depicted in figure 6, we define the OP as an empty NP, ecNP, and put it at the specifier position of the CP. We also define a particular CP, opC2, to represent the CP containing an OP. Notice that the notation "C2" means "C"; that is, the CP.

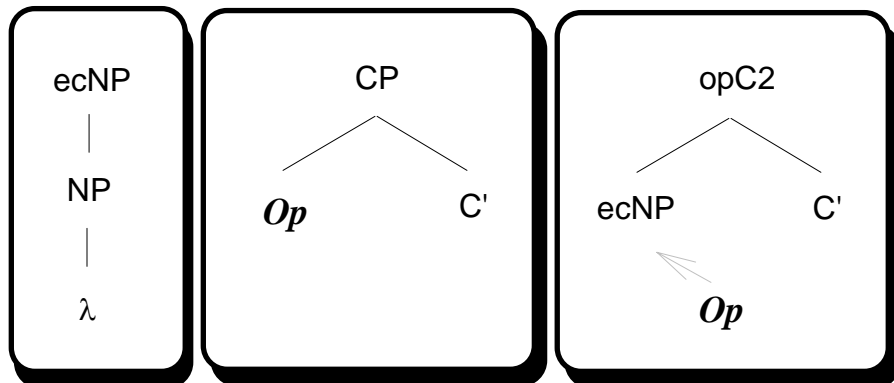


Figure 6: Define the OP as ecNP, and put it at the specifier position of CP (opC2).

```
% To allow Op in [Spec,CP] for relative clauses
```

```
rule ecNP -> [np(NP)] st ec(NP).
```

```
rule opC2$c2 -> [ecNP,c1].
```

```
% To allow Op only in the relative clause context
```

```
rhs [c2,relCINP] replace_rhs [opC2,relCINP].
```

Certainly an adjunction rule which describes the phrase structure that the CP and its adjacent NP can combine together to form an NP is needed. The structure is shown in figure 7:

```
adjunction rule np -> [c2,relCINP] ordered headFinal(n).
```

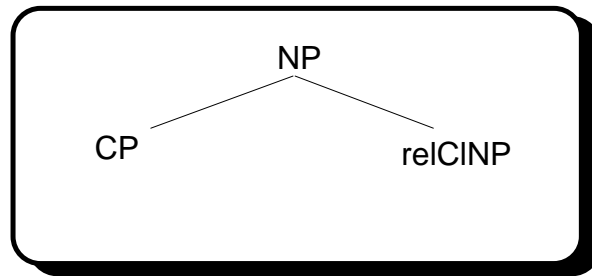


Figure 7: A CP combines its adjacent NP to construct a nominalized structure..

According to the above settings, the relative constructions as depicted in figure 8 will be allowed.

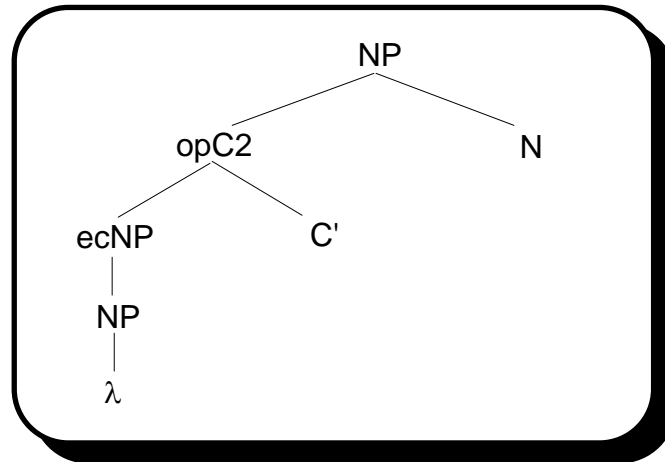


Figure 8: Parse tree for relative clause constructions

4.2. Computational Implementation to Deal with Type-2 Nominalization

As described in section 3.3, the only thing needs to do is encoding the adjunct information. In Pappi, the adjuncts are registered via the predicate "allowExt". For example, if V1 can accept adjuncts with the theta role "instrument"; we can assert the following entry:

`lex('V1',v,[morph('V1',[]),grid([agent],[]), allowExt(instrument)]).`

4.3. Computational Implementation to Deal with Type-3 Nominalization

Listed following are the major part of modifications and implementations for handling type-3a nominalizations: nominalized constituent as a complement of the head noun:

(1) The requirement that only NP needs Case is extended to CP. Thus, the condition that only NP needs inherent Case assignment, .i.e., the statement "cat(CAT,np)", should be eliminated.⁵

⁵There may be some unexpected side effects which affect the other languages; since it's not exactly that CPs in all languages need Case. Hence, in our future work, which constituents need Case will be set as a parameter for the principle.

```

% Change in inherent Case assignment
% Clauses need Case in Chinese
inCaseConfig(CF,Case,CAT) :-
    cat(CAT,np),                                % to be eliminated
    thetaConfig(CF,_,CAT),
    Head head_of CF,
    inherentCaseAssigner(Head,CAT,Case).

```

(2) Also, the highest CP (i.e., *c2 with MatrixLevel(0)*) shouldn't get Case.

```

% Matrix CP shouldn't get Case
rhs [c2(A),mrkr(B)] add_goals [ \+ c_IsMatrixLevel(0)].

```

(3) And then, we encode in the lexicon that "de" is a marker which can realize the genitive case assignment of its left CP:

```

% Realize the genitive Case via "de"
lex('de',mrkr,[left(c2,case(gen),[])]).

```

It means that CP could appear at the left of "de"; and "de" will realize the Case assignment in which the N following "de" will assign genitive Case to the CP.

(4) The language-specific phenomenon in Chinese that the marker "de" can realize genitive Case assignment is declared as:

```

% Define that markers can realize genitive Case in Chinese
realizedAsMarker(gen).

```

Thus, both inherent and structural Cases can be realized by this setting.

```

assignICase(NP,Case,SNP) :-
    .....
assignSCase(Hd,Case,NP) :-
    .....

```

As for type-3b: nominalized constituent as an appositive clause of the head noun, the implementation is quite straightforward: we just need to define an adjunction rule so that CP can be the adjunct of N.

4.4. Computational Implementation to Deal with Type-4 Nominalization

As for the implementation, the only thing needs to do is to modify the following line which declares the constraint that the NP following relative clauses must be overt; in the sense that it can also be an empty NP.

```
rule relCINP -> [overtNP] st true.
```

5. Experimental Results and Conclusions

We evaluated the performance of the Chinese Pappi system by feeding the parser several corpora, and then the parse tree generated by the parser are validated manually. The lexicon file is *automatically converted* from an on-line Chinese dictionary CKIP [Chang et al, 1993] with 81,471 entries. The corpus we use are 127 on-line verdict documents from the Taipei Supreme Court. In each verdict document, one lawsuit case is recorded. There are a total of 16,413 sentences in the corpora with 1,138 sentences containing the word "de". We extracted 379 of them with nominalization for testing.⁶

A simple *pattern matching* mechanism was used to extract the nominalization sentences. The basic consideration is quite obvious: all sentences with nominalizations must contain the keyword "de".

⁶For simplicity, we choose to parse only the sentences containing less than ten words. We will test longer sentences in the future work.

Among these 379 sentences, 356 of them (93.93%) are correctly parsed. Listed in Table 5.1 are example sentences which can be parsed correctly. It's pity that in these judicial documents, type-4 nominalizations don't appear.

Table 5.1. Examples of sentences which can be parsed correctly

| <i>Examples</i> | <i>Nominalization Types</i> |
|---|---|
| 為 [原告 確定 之 論點] <u>is</u> [the plaintiff <u>make sure</u> <u>de</u> <u>argument</u>] It's the argument that the plaintiff made sure. | <i>Relative clauses with complement head nouns</i> |
| 自 無 [禁止 之 理] <u>certainly no</u> [forbid <u>de</u> <u>reason</u>] Certainly there's no reason for forbidding. | <i>Relative clauses with complement adjunct nouns</i> |
| 為 [原告 敗訴 之 判決] <u>is</u> [the plaintiff <u>fail the lawsuit</u> <u>de</u> <u>verdict</u>] It's the verdict that the plaintiff fails the lawsuit. | Nominalization as an appositive clause of the head noun |
| [有 代理權 之 事實] <u>[have agent right de fact]</u> The fact that he has the right of being an agent. | Nominalization as an appositive clause of the head noun |

The major types of sentences which pose outstanding problems to our current theory include: (1) Recursive nominalizations, namely the sentences with the constructions "[CP *de* *Nominalizations*]". (2) Nominalizations with "de" omitted; i.e., the nominalizations with the constructions "[CP *N*]". (3) The "所"-constructions. (4) Sentences contain "倘" (if). Note that type (3) and (4) are general parsing problems that have nothing to do with nominalizations. Examples are listed in Table 5.2.

Table 5.2. Examples of sentences which cannot be parsed well currently

| <i>Examples</i> | <i>Sentence Types</i> |
|--|--|
| <p>被上訴人 主張 之 [前述 之 事項] <u>the defendant claim de [mentioned before de matter]</u> The mentioned matter which the defendant claims.</p> | <i>Recursive nominalizations</i> |
| <p>係 重要 之 [防禦 方法] <u>is important de [defense method]</u> It's an important method of defense.</p> | <i>Nominalizations with "de" omitted</i> |
| <p>能力 <u>所</u> 受 之 損害 <u>ability so suffer de damage</u> The damage that the ability suffers from.</p> | <i>"所"-constructions</i> |
| <p><u>倘</u> 其 權利 之 行使 <u>if his right de performance</u> If the performance of his right,</p> | <i>Sentences contain "倘"</i> |

In this paper, we have introduced how we resolved the nominalization constructions in Mandarin Chinese using the principles and parameters theory, and how we computationalized our analysis by extending Pappi. Experimental results attest to much promise for our system and the theory.

Certainly much work should be done in the future:

- The conjectures that non-matrix CPs require Case and empty NPs can assign Case should be proved under more strict theoretic analysis.
- As discussed in section 3.3, the relative clauses with head nouns as adjuncts should be handled using a better but more sophisticated method.
- As discussed in section 3.4, to avoid unexpected side effects, the kinds of phrases need Case should be parameterized.

- Resolve more linguistic phenomena [Lin & Soo, 1994] in Mandarin Chinese language to make our parser more complete.

Reference

- Aoun, Joseph (1979), *On Government, Case-marking, and Clitic Placement*, mimeographed, MIT, 1979.
- Chang, Li-Ping, et. al (1993), *An Analysis on Chinese Syntactic Categories*, Technical Report no. 93-05, Institute of Information Science, Academia Sinica, Taiwan, ROC., 1993.
- Chen, Hsin-Hsi, I-Peng Lin, and Chien-Ping Wu (1988), *On the Analysis and Interpretation of Chinese Sentences in Logic Programming*, Ph.D. dissertation, Department of Electrical Engineering, National Taiwan University, Taiwan, 1988.
- Chomsky, Noam (1981), *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht; Foris. 1981.
- Chomsky, Noam (1982), *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press, 1982.
- Dai, Jian-Cheng (1994), *A Generalized Unification-based LR Parsing System and Its Application to the Analysis of Chinese Sentences*, Doctoral dissertation, Institutent of Computer Science and Information Engineering, National Chiao Tung University, 1994.
- Fong, Sandiway (1991), *Computational Properties of Principles-based Grammatical Theories*, Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1991.
- Huang, Cheng-Teh James (1982), *Logical Relations in Chinese and the Theory of Grammar*, Doctoral dissertation, Dept. of Linguistics and Philosophy, MIT, 1982.

- Lasnik, H. and J. Uriagereka (1988), *A Course in GB Syntax: Lectures on Binding and Empty Categories*, MIT Press, 1988.
- Li, C. and S Thompson (1981), *Madarin Chinese: a Functional Reference Grammar*, University of California Press, Los Angeles, California, 1981.
- Lin, Dekang (1994), *PRINCIPAR--An Efficient, Broad-coverage, Principle-based Parser*. Proc. of COLING-94, pp.482-488 1994.
- Lin, Koong H.-C. and V.-W. Soo (1994), *Hypothesis Scoring over Theta Grids Information in Parsing Chinese Sentences with Serial Verb Constructions*. Proc. of COLING-94, 1994.
- Lin, Sbin (1995), *Using Principles and Parameters Grammatucal Theories to Parse Mandarin Basic Sentences*, Master Thesis, Dept. of Computer Science, National Tsing-Hua University, Taiwan, 1995.
- Lum, B. and Pun, K. H. (1988), *On Parsing Complex Noun Phrases in a Chinese Sentence*, Proc. of ICCPCOL, pp.470-474, 1988.
- Sells, Peter (1985), *Lectures on Contempory Syntactic Theories: An Introduction to Government-Binding Theory, Generalized Phrase Structure Grammar, and Lexical-Functional Grammar*. Center for the Study of Language and Information, Leland Stanford Junior University. 1985.
- Tang, Tin Chi (1992a), *Studies on Chinese Morphology and Syntax:3*, Taiwan Student Publication Co., 1992.
- Tang, Tin Chi (1992b), *Syntax Theory and Machine Translation: Principle and Parameter Theory*. Proc. of ROCLING V, pp. 53-83, 1992.
- Tsai, Dylan Wei-Ten (1995), *Visibility, Complement Selection and the Case Requirement of CP*. Journal of East Asian Linguistics 4, pp. 281-312.