THE IDENTIFICATION OF THEMATIC ROLES
IN PARSING MANDARIN CHINESE

Keh-jiann Chen*, Chu-Ren Huang**, and Li-ping Chang***

* Institute of Information Science, Academia Sinica
** Institute of History and Philology, Academia Sinica
***Computing Center, Academia Sinica

ABSTRACT

In parsing, the identification of thematic roles not only is fundamental to semantic interpretations but also can reduce syntactic branches and ambiguities. Since the syntactic structures for natural languages are usually complicated and ambiguous, there is no uniquely deterministic parsing strategy. The paper observes that, in addition to real world knowledge, there are four parameters to identify the roles of constituents in a sentence. They are syntactic categories and semantic features of constituents, case frames and case restrictions of verbs, syntactic configurations and word order, and oblique case assigners. Further, the paper shows how the parameters including syntactic and semantic information are represented in feature structures, and how they function in identifying thematic roles. The thematic roles of constituents can be determined by accumulating parametric information. Thus, it is believed that the unification parsing strategy can make parsing more deterministic and efficient.

1. INTRODUCTION

Identification of the thematic role of each constituent in a sentence is a crucial step in the process of understanding natural languages, for well-identified thematic roles reflect the semantic relations between the constituents. In addition to natural languages understanding, thematic roles also play crucial roles in parsing Mandarin Chinese. In Chinese, for lack of morphological markings of derivational variations, there exist many ambiguities in the tree structure of a sentence [陳 88]. Early identification of thematic roles can usually reduce syntactic

123

branches and ambiguities. For example, temporal expressions are instantiated by at least five different syntactic categories as shown in (1). They can be either noun phrases(NPs), compounds with Determinatives and Measures (D-M compounds), postposition phrases(PostPs), preposition phrases(PPs) or adverbs(ADVs)[張 88].

(1)a. *tzaushang* 'morning' / *shiatian* 'summer' -- NPs

  b. *shangshingchi* 'last week' / *houtian* 'the day after tomorrow -- D-M compounds

  c. *kaihuei hou* 'after the meeting' / *shengyi shrbai chian* 'before business fail' -- PostPs

  d. *chen shangke* 'when going to the class' / *tzai gungtzuo shr* 'in working' -- PPs

  e. *tsengjing* 'once' / *yungyuan* 'forever' -- ADVs

If a parser fails to identify any of the above as a temporal expression as early as possible, unnecessary ambiguities may arise. For instance, NPs and D-M compounds could be treated as arguments of a sentence, and the verb in a PostP is likely mis-read as the main verb.

With regard to the definition of thematic roles at senten-tial levels, we generally follow the system of case roles of [Fillmore 68] and thematic roles of [Jackendoff 83]. In addition to arguments, we also take the thematic roles of adjuncts into our consideration. Our system includes modals involving obligation, permission, possibility, etc., and adverbial modifiers of manner, reason, negation, etc. On the other hand, the thematic role at NP level is defined according to the semantic relation between the head noun and the modifier as described in the next paragraph. At other phrasal levels, the semantic relation between the head and a modifier or an argument

is straightforward. For instance, the role that the whole PP plays in a sentence is significant while the relation between the preposition and its object is purely formal.

The thematic roles of arguments and adjuncts represent the semantic relations between them and the head. Such relations can either be very precise or very rough depending on how the thematic roles were defined and assigned. For example, at NP levels , the semantic relations between the modifier and head are very complex, including the relations of predication, quantifier, possessor, apposition and properties. In particular, the relations of properties can be sub-classified into the relation of location, use, time, whole-part, source, shape, etc. However, it is unrealistic for a parser to distinguish them for the implausibility of stacking all the fine-grained real world knowledge to make such trivial distinctions. Hence, our system of thematic roles are stipulated under the premise that a computational identification is plausible. This paper attempts to show how accurate thematic roles can be achieved using syntactically and semantically represented information. In the following section, we will introduce the parameters for the identification of thematic roles. The representations of the grammatical information and their functions in identifying the thematic roles are discussed in the third section. Their actual identification process is proposed in the fourth section.

2. THE PARAMETERS FOR IDENTIFYING THEMATIC ROLES

It is clear that the thematic role of an argument is dependent on its grammatical function and governing verb [Bresnan

82]. However, it does not mean that the determination of the thematic role solely relies on the surface syntactic configurations. For instance, the PP - attachment problem in English as shown in (2), can not be resolved without further information.

(2) John opened the door with the key.

The PP 'with the key' can modify the NP 'the door', or the VP or the whole sentence. However, the most plausible interpretation is that the PP modifies the VP. What, then, are the prime factors to be considered in identifying thematic roles?

In Chinese, thematic roles and syntactic configurations (or word order) also show the intimate correspondences between the thematic role AGENT and the subject position of the active verb, and between the role GOAL and the object position. This is exemplified by *Changsan* as an AGENT in (3)a, and as a GOAL in (3)b.

(3) a. *Changsan da Lisz.*
Changsan beat Lisz
'Changsan beats Lisz.'

b. *Lisz da Changsan.*
Lisz beat Changsan
'Lisz beats Changsan.'

However, for the relatively free word order of Chinese, it is difficult to determine the thematic roles only by the syntactic configurations, as in (4).

(4) a. *Menkou lai le yige ren.*
doorway come Asp a man
'There comes a man at the door.'

a'. *Changsan lai le Taipei.*
Changsan come Asp Taipei
'Changsan came to Taipei.'

126

b. *Chiangshang gua le yifu hua.*
wall   on   hang Asp a   picture
'There hung a picture on the wall.'

b'. *Yifu hua*   *gua*   *chiangshang.*
a   picture hang wall on
'A picture was hung on the wall.'

c. *Jeben shu wo kan guo.*
this book   I read Asp
'The book, I have read.'

c'. *Wo jeben shu kan guo.*
I this book read Asp
'I have read the book.'

Therefore, the determination of thematic roles not only depends on the case frame of the governing verb but also the case restrictions, i.e. a thematic argument must meet the semantic restrictions imposed on that thematic role.

Furthermore, the syntactic category and semantic class(or features) of the constituent also serve as central or partial determining factors in identifying the role of that constituent in the following environments where syntax alone is not sufficient.

a) With Constituents Matching Semantic Restrictions Imposed by the Case Frame of the Verb

(5) *Jeben shu wo kan guo.*
this   book I read Asp
'This book, I have read.'

The verb *kan* 'read' is subcategorized for the case frame<(AGENT)(GOAL)>, and the semantic restriction on AGENT is that it must be animate and the restriction on GOAL is that it must be concrete. Thus in (5) *wo* 'I' has to represent the AGENT and *jeben shu* 'this book' has to represent the GOAL.

127

b) Temporal and Locative Expressions

Since the syntactic categories of these expressions are quite complex, e.g. they can be NPs, PPs, PostPs, ADVs. etc., the thematic role of them can only be determined by their semantic classes.

> (6) *Ta tzuotian mei lai.*
> he yesterday not come
> 'He did not come yesterday.'

As shown in (6), there are two NPs preceding the main verb, the semantic class of *tzuotian* 'yesterday' precludes the possibility of its being an argument of the main verb.

c) Identifying the Roles of Adverbs and Modals

The roles of adverbs are subclassified into manner, degree, quantity, negation, reason, etc. according to their semantic classes. Modals, including (auxiliary) verbs and adverbs, are either subclassified into possibility, obligation, permission, or assigned semantic features according to their meanings.

Both modals and adverbs are sentential adjuncts, so it is easy to identify their thematic roles only from the syntactic categories or their semantic classes, without resorting to syntactic configurations as shown in (7)-(8).

> (7) *Ta hen gaushing.*   ---> DEGREE
> he very happy
> 'He is very happy.'

> (8) *Ni yinggai chiu.*   ---> MODALS,   + obligation
> you should go   'You should go.'

d) Identifying the Thematic Roles of PPs

Generally speaking, the determination of the thematic role of a PP depends on the oblique case assigner (i.e. the

128

preposition) as well as the semantic class of the prepositional object.

> (9) a. *Ta tzai jia kan shu.*
> he at home read book
> 'He read at home.'
>
> b. *Ta tzai jejung chingkung shia kan shu.*
> he in such condition down read book
> 'He read in such condition.'

The thematic role of the PP headed by *tzai* can be time, location, condition, etc. However, the thematic role of the PP *tzai jia* 'at home' in (9)a can only be LOCATION, as determined by the semantic class of the prepositional object *jia* 'home'. Similarly , the PP *tzai jejung chingkung shia* 'in such condition' in (9)b can only have the CONDITION role.

e) Identifying the Thematic Roles of PostPs

In Chinese, a PostP is often composed of a temporal or locative constituent followed by a localizer as shown in (10). Thus, the thematic role of a PostP can only be determined by the semantic type of its argument.

> (10)a. *liangnian yishang* 'over two years' ---> TIME
> two years over
>
> b. *juomian yishang* 'above the desk' --->LOCATION
> desk over

f) Identifying the Semantic Relations Between NP Modifiers and Its Head Noun

The identification of semantic relations between the modifier and the head noun is dependent on the syntactic category and/or the semantic class of the modifier as exemplified by (11).

> (11)a. *neige baubau* 'that baby'
>
> b. *Changsan de baubau* 'Changsan's baby'

c. *Changsan neige* ren   'Changsan, that man'

The semantic relations in (11) are quantifier, possessor, and apposition respectively.

To summarize the above discussion, we conclude that four parameters determine the thematic roles of constituents:

a. the syntactic category and semantic features of the constituent,

b. the case frame and semantic restrictions of the verb,

c. the syntactic configuration and word order, and

d. oblique case assigner, including prepositions and postpositions.

In addition to the parameters, identification of thematic roles sometimes depends also on real world knowledge, as exemplified by the PP in (12) [Winograd 83].

(12) I saw a man on the hill with a telescope.

In Chinese, as mentioned in section 1, the semantic relations between the pre-nominal adjunct of property and the head in an NP depend on real world knowledge, as in (13).

(13) a. *shueiguo mianbau*   --->   MATERIAL
         'fruit bread'

     b. *shueiguo shiaufan*   --->   GOODS
         'the peddler who peddlers the fruit'

     c. *shueiguo pan*   --->   PURPOSE
         'fruit plate'

Making use of real world knowledge can help to identify thematic roles precisely, but the use and the representation of real world knowledge relies on further research in the field of artificial intelligence. Therefore, real world knowledge applied in this paper is limited to the following domain:

a.   the case frame and case restrictions of a verb,

b.   the conceptual hierarchy [Chen 88], in which the semantic features for concepts and the semantic relations between concepts were encoded.

The parameters proposed in this section have been adopted to solve PP-attachments in English, but the actual processes in each adoption are divergent, such as marker - passing in [Hirst 84], two-staged ATN in [林 89], and rule-based approach in [Dahlgren 86].   There is yet to be a thorough discussion on the identification of the thematic roles in Chinese.   In the next section, we will propose the idea for the process and give the implementation detail in the section 4.

3. REPRESENTATION AND UNIFICATION OF PARAMETRIC INFORMATION

Syntactic structures for natural languages are usually complicated and ambiguous.   There is no uniquely deterministic parsing strategy for natural languages.   Marcus[80] considered deterministic parsing for natural languages is possible, although only under the condition that the parser can look three constituents ahead and disregard ambiguities such as PP-attachment. In other words, in syntax-only parsing, backtracking is unavoidable without looking ahead [Marcus 80].   However, language comprehension of human beings does not seem to involve looking ahead nor backtracking.   In other words, identification of thematic roles can be done left to right without the hypothesis-correcting process of looking ahead or backtracking. It is reasonable to assume that this is because that syntactic

information, semantic information as well as real world knowledge are involved in human comprehension. From a communicative point of view, human languages should avoid ambiguities and nondeterminism in order to reduce misunderstanding. Therefore we believe that human language comprehension can be done by simply accumulating information monotonously. Violations of the above principle are rare but do appear in some languages. Garden-path sentences present typical examples as in (14) [Hirst 84].

(14)a. The horse raced past the barn fell.

b. The old dog the foot steps of the young.

However, we fail to discover similar examples in Chinese. Hence we believe that the identification of thematic roles in this language can be done by way of information accumulation with simple inferences and reasoning. The feature structures and unification formalisms advocated by many current theorists [Bresnan 82, Gazdar 85,87, Kay 85, Pollard 87, Shieber 86, Uszkoreit 86] offer a good approach to encode and accumulate the information needed in identifying thematic roles. Accordingly, we propose to unify syntactic analysis and semantic interpretation into a single process. Thus not only full advantages of unification parsing strategy are taken but both syntactic and semantic information also complement each other to make parsing more deterministic and efficient. In the next section, we will take Chinese examples to show how the parameters are represented and how they function in identifying the thematic roles.

4. THE IDENTIFICATION OF THEMATIC ROLES

As described in section 2, in addition to real world knowledge, there are four parameters to identify thematic roles of constituents in a sentence. The representation for the parametric information is specified in the lexical feature structure of the Information-based Case Grammar(ICG)[陳 89] . The information of the four parameters includes the syntactic information, like syntactic category, and word order, etc., and semantic information, like case restrictions, semantic features, etc. ICG offers lexicon-based syntactic representations. Each lexical entry contains the syntactic and semantic feature structures for the lexical head and projection phrases, as denoted in (15)*.

```
(15)   word : [ sem [ meaning:
                      features:
                      arguments:
                      adjuncts:
               syn [ class:
                     constraints: [ Form:
                                     BP (Basic Pattern):
                                     AP (Adjunct Precedence):
```

Let us recall that the first parameter in identifying thematic roles involves syntactic categories and semantic features. The information is specified in the syntactic class column and in the semantic features column respectively, as shown in (16)-(18).

```
(16)   jungchiou 'Mid-Autumn':
              [ sem [ features:  + time
                      ...
                syn [ class:  Nd (temporal noun)
                      ...
```

---

* For convenience and readability, only partial information of the description of the information structure is given in this paper. Readers are referred to [陳 89] for detailed description and discussion of the grammar formalism.

(17) *keneng* 'possibly':

$$\begin{bmatrix} \text{sem} \begin{bmatrix} \text{features:} & + \text{modal, } + \text{possibility} \\ & \cdots \end{bmatrix} \\ \text{syn} \begin{bmatrix} \text{class:} & \text{VL3 (modality)} \\ & \cdots \end{bmatrix} \end{bmatrix}$$

(18) *kungshou* 'empty-handed':

$$\begin{bmatrix} \text{sem} \begin{bmatrix} \text{features:} & + \text{manner} \\ & \cdots \end{bmatrix} \\ \text{syn} \begin{bmatrix} \text{class:} & \text{Dh (ADV)} \end{bmatrix} \end{bmatrix}$$

The second parameter involves case frames and case restrictions of verbs. The information is specified as the values of the semantic arguments, as in (19)-(20).

(19) *da* 'beat':

$$\text{sem} \begin{bmatrix} \text{arguments:} \begin{bmatrix} \text{AGENT:} [\text{features:} & + \text{animate} \\ \text{GOAL:} [\text{features:} & + \text{physical} \end{bmatrix} \end{bmatrix}$$

(20) *chiuan* 'persuade':

$$\text{sem} \begin{bmatrix} \text{arguments:} \begin{bmatrix} \text{AGENT:} [\text{features:} & + \text{human} \\ \text{GOAL } \boxed{1} \text{ *:} [\text{features:} & + \text{human} \\ \text{THEME:} \begin{bmatrix} \text{feature:} & + \text{active} \\ \text{arguments:} & \text{AGENT: } \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The third parameter involves the syntactic configuration and word order. In ICG, the syntactic rules of the phrasal and the sentential structure can be represented in the information structure of the head by means of the immediate dominance rules(ID rules) and linear precedence rules(LP rules). The immediate dominance relations are specified in the semantic information column encoding the argument specification and the adjunct specification, and the syntactic form and the linear precedence of the constituent are specified in part of the syntactic constraints column [陳 89], as shown in (21)-(22).

---

* ' 1 ' indicates the co-referential relation, i.e. the GOAL and the AGENT of the embedded sentence are identical.

(21) *chiuan* 'persuade':
```
 ┌sem: ┌arguments: [AGENT, GOAL, THEME
 │     └adjuncts:  time, loc, modality, reason,
 │                     instrument, manner, ...
 │syn: ┌constraints: ┌form: ┌time[{NP, DM, PP, ADV, +time}]
 │     │             │      │loc[PP, +location]
 │     │             │      │modality[VL3]
 │     │             │      └manner[ADV, +manner]
 │     │             │ BP:  AGENT[NP]< * <GOAL[NP]<THEME[VP]
 │     │             │ AP: ┌1. {modality, time, loc} < *
 └     └             └     └2. AGENT < manner < *
```

(22) *da* 'beat':
```
 ┌sem: ┌arguments: [AGENT, GOAL
 │     │adjuncts:  time, loc, modality, manner, reason,
 │     └               instrument, ...
 │syn: ┌form:  (omitted)
 │     │BP: ┌1. AGENT[NP] < * < GOAL[NP]
 │     │    └2. GOAL[NP] < AGENT[PP[bei]] < *
 └     └AP:  (omitted)
```

The last parameter is case markers. The oblique case assigners generally are represented as prepositions in Chinese, for example, *ba* is a GOAL marker, *bei* is an AGENT marker. However, a preposition may mark more than one case depending on the object of the preposition[黃 88]. For example, when the object of the preposition *bei* is an animate NP, *bei* in (23) marks an AGENT case; when the object is an abstract NP, *bei* in (24) marks a CAUSER case; when the object is a non-animate NP, *bei* in (25) marks an INSTRUMENT case.

(23) *Changsan bei Lisz da.*
     Changsan BEI Lisz beat
     'Changsan was beaten by Lisz.'

(24) *Changsan bei emeng jingshing.*
     Changsan BEI nightmare awake
     'Changsan was awakened by nightmare.'

(25) *Changsan bei jen tszpuo le shoujr.*
     Changsan BEI needle sting Asp finger
     'Changsan was stung on the finger.'

In ICG, the case assigned by the preposition is represented

as semantic features in the information structure, exemplified by *bei* in (26).

```
(26)  bei:
   ⌐ sem ⌐ features:  1.AGENT   2.CAUSER   3.INSTRUMENT
   |      |  arguments: DUMMY:  ⌐ features: 1. +animate
   |      |                     |             2. -physical
   |      └                     └             3. -animate
   | syn ⌐ form:   DUMMY[NP]
   └     └ BP:   *  << DUMMY
```

As shown in (26), the No.1 semantic feature has to correspond with the No.1 DUMMY feature, and so on so forth. In other words, when *bei* marks an AGENT, the object of *bei* must be animate. Other predictions follow similarly.

The thematic roles of the PPs in (23)-(25) is fully identified by the preposition *bei* and its object. But the identification of thematic roles for certain other PP's can not be completely resolved by the case marker and its object, since it still depends on the information of verbs. Further discussion of this issue will be given later involving (32)-(35).

With the representation of the four parameters in identifying thematic roles described, we can classify five different cases of how the parameters work in application.

4.1. *IDENTIFIED ONLY BY THE FIRST PARAMETER -- THE SYNTACTIC CATEGORY AND THE SEMANTIC FEATURE OF THE CONSTITUENT*

a. Adverbs: the thematic roles of adverbs are exhaustively and unambiguously classified into quantity, evaluation, negation, degree, manner, reason, time, location, etc.[張 89], according to their semantic class.

b. Modals: modals are classified into obligation, permission, possibility, etc.[張 89] according to their semantic feature.

c. A small part of NPs: the identification of the thematic roles of most NPs is related with case restrictions, case frames, even with prepositions. However, there are some special types of NPs which are identified only by semantic features, for instance, time NPs.

For the above categories, their thematic roles can be immediately identified, so many unnecessary syntactic branches and ambiguities can be avoided.

4.2. IDENTIFIED BY TWO PARAMETERS -- THE CONSTITUENT AND THE OBLIQUE CASE ASSIGNER

a. PPs: the uncertainty in identifying thematic roles of PPs can often be resolved by the semantic features of the prepositional object as exemplified by (27)-(29).

(27) *bei Lisz* 'by Lisz':
    sem ⎡ features: AGENT,   +human
        ⎣ arguments: DUMMY:⎡meaning: Lisz
                             ⎣features: +animate
(28) *bei emeng* 'by nightmare':
    sem ⎡ features: CAUSER, +phenomenon
        ⎣ arguments: DUMMY: ⎡meaning: nightmare
                              ⎣features: -physical
(29) *bei jen* 'by the needle':
    sem ⎡ features: INSTRUMENT, +artificiality
        ⎣ arguments: DUMMY: ⎡meaning: needle
                              ⎣features: -animate

Based on the conceptual hierarchy [Chen 88], human inherits the feature animate, so according to (26) the feature structure of *bei*, *bei Lisz* 'by Lisz' in (27) is identified with the No.1 role AGENT. The thematic roles of PPs in (28) and (29) are identified in the same way. However, for certain prepositional phrases, their resolution of thematic roles cannot be uniquely determined by two parameters, the uncertain thematic roles will be further resolved at

sentential level. We will discuss them in 4.3.

b. PostPs: a postposition in Chinese often marks TIME or LOCATION, and the constituent preceding a postposition may neither be temporal nor locative, as shown in (30).

(30) a. _shueichr_ *chian*   'before the water pond'

   b. _chushr_ *chian*   'before things-go-wrong'

Then, how can one identify the thematic role of the PostP based on the semantic features of the pre-argument of the postposition?   Take the postposition *chian* 'before' as example, its feature structure is as follows.

(31) *chian* 'before':

```
  sem [ features:   1.TIME      2. LOCATION
      [ arguments:   DUMMY: [ features: 1. {+ time, + event}
                                        2. + physical
  syn [ form:  DUMMY[{NP, VP, S}]
      [ BP:    DUMMY << *
```

Since *shueichr* 'pool' in (30)a has the feature physical, so the thematic role of the PostP in (30)a can only be LOCATION, as stipulated by the feature specifications. By the same token, the syntactic category and the semantic feature of *chushr* 'things-go-wrong' in (30)b matches the No.1 semantic feature of the post-position *chian* 'before', so the thematic role of the PostP in (30)b is TIME.   Thus, the thematic roles of all PostPs can be best identified at the phrasal level.

4.3. *IDENTIFIED BY THREE PARAMETERS -- THE CONSTITUENT,*
   CASE FRAME AND CASE RESTRICTION, AND OBLIQUE CASE ASSIGNER

In this section, we will focus on the unresolved thematic roles of some PPs after the phrasal level parsing. These will be determined by three parameters after parsing the sentence, typical examples are given in (32).

(32)a. *Shiuesheng yi wenchiuan shitzau.*
    student   YI hot spring bathe
    'The students bathe in the hot spring.'

b. *Beitou yi wenchiuan wenming.*
   Beitou YI hot spring famous
   'Beitou is famous for hot springs.'

c. *Yi wenchiuan eyan, Yangmingshan tzueihau.*
   YI hot spring   the best
   'As for hot springs, the Yangming mountain offers the best.'

Though the three PPs are headed by the same preposition *yi*, their thematic roles are different owing to the different contextual environments, The PP is an INSTRUMENT in (32)a, a CAUSER in (32)b, and a TOPIC in (32)c.

The thematic roles of the PPs in (32) can not be adequately determined only by the feature structure information of *yi* in (33).

(33) *yi*:
```
 ┌ sem ┌ features:  1.INSTRUMENT  2.MANNER  3.CAUSER  4.TOPIC
 │     │ arguments:  DUMMY: ┌ features:  1. -animate
 │     │                    │            2. {-physical, +event}
 │     │                    │            3. {+entity, +event}
 │     └                    └            4. {+entity, +event}
   syn ┌ form:  1. DUMMY[NP], 2.3.4. DUMMY[{NP,VP,S}]
 ╲     │ BP:    ┌ B1:  1.2.3.* << DUMMY
       │        ⟨ B2:  4.* << DUMMY << {laishuo, laikan, shuolai,
       └        └      tinglai, kanlai, eyan, eluen}
```

After parsing the PP *yi wenchiuan* in (32), the unified information shows the following result:

(34)   *yi wenchiuan* 'with the hot spring'
```
 ┌ sem ┌ meaning:  hot spring
 │     │ features:  1.INSTRUMENT  3.CAUSER
 │     └ arguments:  DUMMY: [ features:  terrain
 └ syn [ class:  PP
```

MANNER and TOPIC are out of consideration respectively for conflicts in semantic features and the BP of the PP. With regard to the distinction between INSTRUMENT and CAUSER, it depends on further information to be supplied by the main verb. If the verb

is active, then the thematic role of the PP is an INSTRUMENT; if the verb is stative, the thematic role is a CAUSER. The parameters functioning in identifying the thematic role of the PP headed by *yi* are as the following:

(35)

case marker	syntactic & semantic restrictions of the constituent	class of verb		role
yi	- animate	active		INSTRUMENT
	- physical or + event	active		MANNER
	having the following constituent{laishuo, eyan, ...}	active, stative		TOPIC
	+event	stative		CAUSER

Although the above discussion seems a bit complicated, in fact, the process is very simple when the related deterministic information is represented at different levels. For example, the information structure of the *yi* PP unifies the information of the preposition *yi* in (33) and the semantic information of its possible argument to get a set of the unresolved thematic roles as shown in (34). The additional information needed is represented in the information structure of the verb, as the following rough representation:

(36) active verbs:
syntactic   [form: [INSTRUMENT[PP[{*yung, yi, kau, jie, ...*}]]
information: MANNER[PP[{*yi, yung, ping, kau, ...*}]]
                         CAUSER[PP[{*bei, jiau, rang, wei, ...*}]]]

(37) stative verbs:
syntactic
information: [form: [MANNER[PP[{*yi, yung, jie, jang, ...*}]]
                    CAUSER[PP[{*yi, bei, jiau, wei, ...*}]]]

In (37), the INSTRUMENT case role is not specified in the information structure of stative verbs; In (36), *yi* is not

involved in the CAUSER. Further, when the information of the preposition *yi* in (33) and the stative verb *wenming* 'famous' unify, the thematic role of the PP in (32)b can only be a CAUSER because it is the only possible role stipulated by the verb. On the other hand, when the information of the PP unifies with the information of the active verb *sitzau* 'bathe', the thematic role of the PP in (32)a has to be an INSTRUMENT because that is the only possible one specified by the verb.

4.4. IDENTIFIED BY THREE PARAMETERS -- THE CONSTITUENT, THE VERB AND WORD ORDER

The thematic roles of most arguments unmarked by case assigners are identified by three parameters.

(38) *Jeben shu wo kan guo.*
this book I read Asp
'This book, I have read.'

In (38), *Jeben shu* 'the book' is the GOAL of the verb *kan* 'read', and *wo* 'I' is the AGENT. The related parametric information is encoded in the information structure of the verb *kan*, as shown in (39).

(39) *kan* 'read':
```
⌈ sem ⌈ arguments: ⌈ AGENT:  features:  + animate
│       ⌊             ⌊ GOAL:   features:  + physical
│ syn ⌈ BP: ⌈ B1:  AGENT < * < GOAL
│      │    ⟨ B2:  AGENT < GOAL[+definite] < *
⌊      ⌊    ⌊ B3:  GOAL[+definite] < AGENT < *
```

In parsing, functional identification of the thematic roles is unresolved after parsing the NPs *jeben shu* 'the book' and *wo* 'I'. Their thematic roles are not resolved until the unification of the information of the verb *kan* 'read'.

4.5. IDENTIFIED BY ALL FOUR PARAMETERS

(40) *Changsan bei Lisz da.*
Changsan BEI Lisz beat
'Changsan was beaten by Lisz.'

In (40), the thematic role of the NP *Changsan* is identified with the information of the constituent, the case frames and restrictions of the verb, and word order. In addition, it is also related to the case role of the PP *bei Lisz*, for the preposition marks the AGENT case. Based on the information structure of the verb *da* 'beat', given in (22), it requires a GOAL and an AGENT, and one possible syntactic encoding has the GOAL occur preceding the *bei*-marked AGENT. Thus the four parameters converge to determine the thematic role of the NP *Changsan* as a GOAL.

The actual examples described in section 4 roughly show how the parametric information is represented and how it functions in identifying thematic roles at different phrasal level. Information accumulation is a general method to identify thematic roles. It is not only applicable to the processing of Chinese. For instance, we think the PP - attachment problem in English also can be resolved by the same way.

With regard to real world knowledge, it has not been implemented for representational difficulties. However, the concept of information accumulation is also applicable to the processing of real world knowledge when available.

5. CONCLUSION

In parsing, the identification of thematic roles is a very important procedure. It not only is fundamental to semantic interpretations but also reduces syntactically ambiguous branches

142

if performed timely.

The identification of thematic roles is just a start point to semantic interpretation and to the understanding of natural languages. Some of the thematic roles discussed are still very crudely defined owing to the limitation of applicable information. The available information only includes a) syntactic categories and semantic features of the constituent, b) the case frame and case restrictions of the verb, c) the syntactic configurations and word order, and d) the oblique case assigners. Correct identification of thematic roles depends on how much information is available. The more available information is, the better our identification should be. Even though the roles we have chosen to use are roughly those proposed by Jackendoff[83], we did make some refinement. Furthermore, if real world knowledge is better represented in the future, the thematic role can be more precisely and quickly identified.

REFERENCE

林雅萍　1989，"嵌入 X' 理論於 ATN 系統中以解決介詞片語所屬問題"，碩士論文．
台北：台灣大學．

陳克健　1988，"中文剖析的問題與對策"，中華民國第一屆計算語言學研討會論文集．
pp. 21-28， 南港：中央研究院計算中心．

陳克健，黃居仁　1989， "訊息為本的格位語法 -- 一個適用於表達中文的語法模式"，
第二屆計算語言學研討會論文集．

張莉萍　1988，"漢語的時間詞組和語言剖析"，中華民國第一屆計算語言學研討會
論文集． pp. 75-86，南港：中央研究院計算中心．

張麗麗及中文詞知識庫小組　1989，國語的詞類分析， 修訂版， 中央研究院計算中心
技術報告．

黃瑞珠　1988，"由剖析的觀點分析漢語介詞組"，中華民國第一屆計算語言學研討會
論文集． pp. 127-144，南港：中央研究院計算中心．

Bresnan, J. (ed.) 1982. *The Mental Representation of Grammatical Relations.* Cambridge, Mass. : MIT Press.

Chen, K. J. and C. S. Cha 1988. "The Design of a Conceptual Structure and Its Relation to the Parsing of Chinese Sentences." ICCPCOL'88, Toronto.

Dahlgren, K. and J. McDowell 1986. "Using Commonsense Knowledge to Disambiguate Prepositional Phrase Modifiers." AAAI-86, 589-593.

Fillmore, C. 1968. "The Case for Case." In E. Bach and R. Harms (Eds.), Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston.

Gazdar, G., E. Klein, G. K. Pullum, and I. A. Sag 1985. *Generalized Phrase Structure Grammar.* Cambridge: Blackwell, and Cambridge, Mass.: Harvard University Press.

Gazdar, G. et al. 1987. *Category Structures.* CSLI Report 102, Stanford: CSLI.

Hirst, G. J. 1984. "Semantic Interpretation Against Ambiguity."
    Ph.D. Diss. Brown University.

Jackendoff, R. 1983. *Semantics and Cognition*. MIT Press.

Kay, M. 1985. "Parsing in Functional Unification Grammar." In
    D. Dowty, L. Karttunen, and A. Ziwicky (Eds.), Natural
    Language Parsing. Cambridge: Cambridge University Press.

Marcus, M. P. 1980. *A Theory of Syntactic Recognition for
    Natural Languages*. MIT Press.

Pollard, C. and Ivan A. Sag 1987. *Information-based Syntax
    and Semantics, Vol.1 Fundamentals,* CSLI Lecture Notes
    Series No. 13. Stanford: CSLI.

Shieber, S. 1986. *An Introduction to Unification-Based
    Approaches to Grammar*. CSLI Lecture Notes Series No.4.
    Stanford: CSLI.

Uszkoreit, H. 1986. "Categorial Unification Grammars." In
    Coling 1986. Bonn: University of Bonn. Also appeared as
    Report No. CSLI-86-66, Stanford: CSLI.

Winograd, T. 1983. *Language as a Cognitive Processes*. Vol.1,
    Syntax, Addison-Wesley.