

# 中文句結構樹資料庫的構建 (Sinica Treebank)

陳鳳儀\*, 蔡碧芳\*, 陳克健\*, 黃居仁<sup>+</sup>  
Feng-Yi Chen\*, Pi-Fang Tsai\*, Keh-Jiann Chen\*, Chu-Ren Huang<sup>+</sup>

## 摘要

中文句結構樹資料庫 (Sinica Treebank) 建構的主要目的是提供中文自然語言處理研究一個具有標記語料庫的研究素材, 我們可以從這個中文句結構樹資料庫中抽取語法知識, 也藉由語法知識的抽取與瞭解使我們的剖析系統功能更趨完善。本文介紹中文句結構樹資料庫構建方法和步驟, 從五百萬詞的中央研究院平衡語料庫 (Sinica Corpus), 抽取句子, 以訊息為本格位語法 (Information - based Case Grammar, ICG) 的表達模式為基本架構, 經由電腦自動剖析成結構樹, 可以盡量維持結構標記的一致性, 最後並加以人工修正、檢驗, 以維持標記的正確性。對於歧義的句法結構形式及詞類標記, 我們也提出處理的原則。

## 1. 緒論

中文句結構樹資料庫 (Sinica Treebank) 是中央研究院詞庫小組從具有標記的五百萬詞的中央研究院平衡語料庫 (Sinica Corpus) 中, 抽取句子, 經由電腦剖析成結構樹, 並加以人工修正、檢驗後的所得的成果。中文句結構樹資料庫建構的主要目的是提供中文自然語言處理研究一個具有標記語料庫的研究素材, 我們可以從這個中文句結構樹資料庫中抽取語法知識, 也藉由語法知識的抽取與瞭解使我們的剖析系統功能更趨完善。

本文的目的是介紹中文句結構樹資料庫構建方法和步驟。為了有效率並且精確的剖析中文句子, 我們提出以訊息為本格位語法 (Information - based Case Grammar, ICG) 的表達模式。利用這個兼顧語法和語意訊息的表達方式, 每一個中文句結構樹既有語法的結構分析, 還顯示出每個詞之間的語意關係。我們將在第二節介紹訊息為本的格位語法 (ICG) 表達模式。

---

\* 中央研究院資訊所, 陳鳳儀: apple@iis.sinica.edu.tw 蔡碧芳: panhu@iis.sinica.edu.tw  
陳克健: kchen@iis.sinica.edu.tw

<sup>+</sup> 中央研究院語言所, 黃居仁: hschuren@ccvax.sinica.edu.tw

在第三節，中文句結構樹資料庫的構建方法和步驟，可以使我們的分析更一致。在第四節，針對特殊結構，例如並列結構、主題句等一一提出標記和分析原則。最後一節是結論。

## 2. 訊息為本的格位語法表達模式

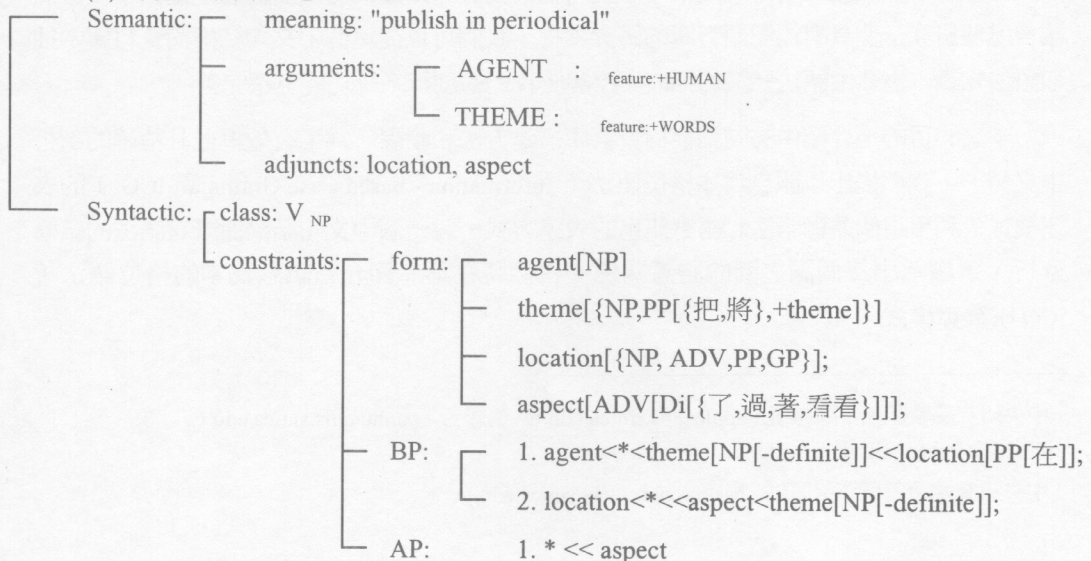
為了有效率並且精確的剖析中文句子，我們提出訊息為本格位語法（Information - based Case Grammar, ICG）的表達模式。這個以詞彙為中心，並且配合「中心語主導原則」（Head-Driven Principle）的中文剖析系統，對於剖析的句子除了記載語法訊息外，也藉由語意角色的指派標示出詞和詞之間的語意限制關係 [ 陳克健 1989, Chen 1992, 1996, Chen et al 1994, 1999 ]。

### 2.1 詞彙訊息的表達

在 ICG 架構下，每一個詞彙都記載了語法和語意兩種訊息。在語意訊息方面，不但記載了意義 (meaning/features)，還包含其支配的論元 (arguments) 和可能的修飾成分 (adjuncts); 在語法訊息方面，記載了語法類別 (class) 和其語法限制 (constraints)，語法限制又包含了型式上的限制，如成分形式 (form) 和詞序上的限制，如表示中心語和論元可能組合的基本句型 (BP)、表示成分之間順序關係的成分規律 (AP) [ 陳克健 1989, Chen 1992, 1996, Chen et al 1994, 1999 ]。

以「刊登」為例，在語意訊息方面，它的意義我們暫時以英文翻譯來表達，而論元為主事者 (agent) 和客體 (theme)，可能的修飾成分有地點 (location)、時態標記 (aspect) 等；在語法訊息方面，「刊登」的語法類別為動作單賓動詞，無論是論元或修飾成分，每一個成分都分別記載了語法形式，而基本句型和成分規律皆以線性規律表達相關的詞序關係。

#### (1) "刊登" 詞彙訊息的表達方式



## 2.2 中心語主導原則

中文句子的語法結構常常有歧義的現象。如果只是表達詞和詞的語法結構，不足以解決中文這種相當程度的歧義現象。因此為了解決這種歧義性，我們採取中心語主導原則 (Head-Driven Principle)。中心語主導原則是指每一個句子或詞組結構皆有一中心語 (Head)，詞組結構由中心語與其論元 (argument) 或附加成分 (adjunct) 組成，並由中心語決定詞組的詞類，例如：句子 (S) 和述詞詞組 (VP) 的中心語皆為述詞 (V)；名詞詞組 (NP) 的中心語為名詞 (N)；介詞詞組 (PP) 的中心語為介詞 (P)；方位詞詞組 (GP) 的中心語為方位詞 (Ng)；對等連接詞 (XP) 的中心語為連接詞 (C)。藉由中心語指導原則，在剖析中文句子的時候，不但可以決定每一個中心語的詞組類型，並且利用中心語和其他成分所記載的語法和語意訊息，將句子中詞和詞之間的語法和語意的限制關係清楚的表達出來 [陳克健 1989, Chen 1992, 1996, Chen *et al* 1994, 1999]。

以例句 (2a) 和 (2b) 說明如下：動詞「刊登」為 (2a) 的中心語，根據 (1) 所記載的訊息決定了句中每一個成分的結構，並同時指派其語意角色，「他」為「刊登」的主事者 (agent)，語法形式為名詞組 (NP)，「一則廣告」為「刊登」的客體 (theme)，語法形式為包含一定量詞組 (DM) 的名詞組，而「在報紙上」是由表地點 (location) 的介詞「在」所引介的介詞組 (PP)，句子剖析的結果如 (2b)。

(2a) 他刊登一則廣告在報紙上。

(2b) S (agent: NP (Head: N<sub>haa</sub>: 他) | Head: V<sub>NP</sub>: 刊登 | theme: NP (quantifier: DM: 一則 | Head: N<sub>ac</sub>: 廣告) | location: PP (Head: P<sub>21</sub>: 在 | DUMMY: GP (DUMMY: NP (Head: N<sub>ab</sub>: 報紙) | Head: Ng: 上)))

## 3. 中文句結構樹建構步驟

為了使中文句結構樹資料庫的構建有效率並且一致，我們首先從具有標記的五百萬詞的中央研究院平衡語料庫 (Sinica Corpus) 中，抽取句子，經由以 ICG 為基本架構的中文剖析系統，自動剖析並且產生結構樹，再利用句結構樹編輯程式並配合標記原則，加以人工修正和檢驗。目前已經完成了 38,725 棵中文結構樹，共有 239,532 個詞，而所選取的文章主題涵蓋政治、旅遊、運動、財經、社會等 [Chen *et al* 1996]。

### 3.1 中文句結構樹的詞組結構和標記原則

我們的中文剖析系統在剖析的過程中，依據記載於詞彙中的語法、語意訊息和中心語主導原則，決定每一字串的中心語的詞組類型。當中心語的詞組類型決定時，每一個詞組結構都標上類型和相關的語意角色。因此每一個成分都標記有語意角色、詞 / 詞組類型等。以下舉例說明，例句 (3a) 經過電腦剖析產生的結構樹為 (3b)，其中每一個成分皆以 "|" 分



開：

(3a) 那個挽髻的女人白髮之後便不再理會庭前亭亭玉立的青草

(3b) #S(agent:NP (quantifier:DM: 那個 | property:VP • 的 (head:VP (Head:VA4: 挽髻) | Head:DE: 的) | Head:Nab: 女人) | time: GP (DUMMY: VP (Head: VH11: 白髮) | Head: Ng: 之後) | time: Dd: 便 | time: Dd: 不再 | Head: VC2: 理會 | goal: NP (property: VP • 的 (head: VP (location: NP (property: Ncb: 庭 | Head: Ncda: 前) | Head: VH11: 亭亭玉立) | Head: DE: 的) | Head: Nab: 青草)) #

(4) 符號說明：

#：以“#”置於前後，做為一結構樹段落。

()：詞組的組合成分為複雜結構，以“(、)”標示其詞組結構左右邊界。

|：分隔在同一層次上的成分結構。

在例句(3b)中，「理會」為整個句子的中心語(Head)，它決定了二個必要論元：agent「那個挽髻的女人」和goal「庭前亭亭玉立的青草」；time「白髮之後」、time「便」、time「不再」則為附加成份。

"agent: NP: 那個挽髻的女人"結構中，中心語是"Head:Nab: 女人"，它包含了二個修飾語，一是"quantifier:DM: 那個"，一是"property:VP • 的 (head:VP(Head:VA4: 挽髻) | Head:DE: 的)"。同理"property:VP • 的 (head:VP(Head:VA4: 挽髻) | Head:DE: 的)"是由兩個組合成分組成的，其結構為一個語義上的中心語"head:VP(Head:VA4: 挽髻)"和一個語法上的中心語"Head:DE: 的"。以下的time, location和goal依此類推。

而語意角色的說明和詞組結構類型的標記與標示的原則分別簡述如下，至於詞類的標記和語意角色的詳細內容，請參見附錄或詞庫小組技術報告CKIP-93-05[詞庫小組1993]和CKIP-92-01[林甫雯1992]。

(5) 語意角色：

Head：語法上的中心語。表示句子或任何詞組結構的中心語。

head：語義上的中心語。中文有些結構，如「VP • 的」形式，其中心語我們不容易決定，因為在形式上它的中心語為"的"，但實際上，它的語意是由「的」前面的VP所決定；因此，我們以Head表示形式上也就是語法上的中心語，以head表示實際載有意義的語義中心語。

(6) 詞組結構：

S：表示結構樹為句子(S)；以述詞為中心語；此外當主詞和述詞的賓語或



補語的型式為句子或子句的時候，詞組結構標記為 S，不為 NP。

- a. VP：述詞詞組；中心語為述詞（V）。
- NP：名詞詞組；中心語為名詞（N）。
- GP：方位詞詞組；中心語為方位詞（Ng）；所帶論元角色為 DUMMY<sup>1</sup>。
- PP：介詞詞組；中心語為介詞（P）；所帶論元角色亦為 DUMMY。
- XP：連接詞詞組；中心語為連接詞（C）；X 代表一個變數，XP 的真正詞類由連接成分決定，例如：連接成分為述詞詞組（VP），則為述詞詞組（VP），連接成分為名詞詞組，則為名詞詞組（NP）。
- DM：定量詞詞組。

除了以上的詞組結構以外，還有其他的詞組結構值得加以說明：(1){的，地，之，得，到，個}等可以分別與任何的詞組再組合成另一個詞組，其表達方式為 {A，N，V，S，DM，GP，NP，PP，VP，ADV} • {的，地，之} 或 {的，地，得，到，個} • {NP，VP，S}。其中，在我們的表達結構中，我們以 A • {的，地，之} 來表達形容詞詞組結構，而不是以 AP 來表達；同理，我們以 ADV • {的，地，之} 來表達副詞詞組結構，而不是以 ADVP 來表達；(2)N 或 V 也表示詞組的標記，當 N 或 V 的語意角色指派以後，其上層結構才標上 NP 或 VP。以 N 或 V 表示詞組的使用限制將在第四節有詳細的討論。

### 3.2 人工編修和檢測

中文字串經過電腦自動剖析可能產生一個到數百個結果，甚至有時可能無法產生任何結果，所以電腦自動剖析後的結果還需要人工來編輯或篩選出最好的剖析句子。為了使人工編輯的結果一致，並且有依循的根據，我們將在第四節提出中文句結構樹構建的輔助原則。另外為了使人工編輯更有效率，我們完成了中文句結構樹資料庫的編輯程式。這個編輯程式的主要功能有 [Chen *et al* 1999]：

- A. 可以自動將線性結構轉換成樹狀結構圖並且顯示出來，以方便編輯人員修改或選取。再以 (3a) 和 (3b) 為例，(3b) 是電腦自動剖析 (3a) 後的結果，以線性結構的方式表達，而下列的例句 (7) 為 (3b) 的樹狀結構圖。

<sup>1</sup> 我們規定方位詞和介詞所帶論元的語意角色為 DUMMY，表示一個未定的語意角色。方位詞或介詞所帶論元的語意角色不似動詞是固定不變的，必須根據實際語境來指派適當的角色，如例 (1) 和例 (2) 中，名詞組「藍絲綢襯衫」和「政治」在以「上」為中心語的方位詞組 (GP) 中，其語意角色分別為 "location" 和 "condition"。

(1) VP(location:GP(DUMMY:NP(property:VH11: 藍 |property: Naa: 絲綢 | Head: Nab: 襯衫)| Head: Ng: 上)| Head: VC33: 釘滿 | goal: NP (property: Nac: 圓形 | Head: Nab: 亮片))

(2) VP (condition: GP ( DUMMY:NP(Head:Nad: 政治)|Head:Ng: 上 )|deontics:Dbab: 不能 |evaluation: 真正 | Head: VC2: 實行 | goal: NP ( Head: Nad: 民主))

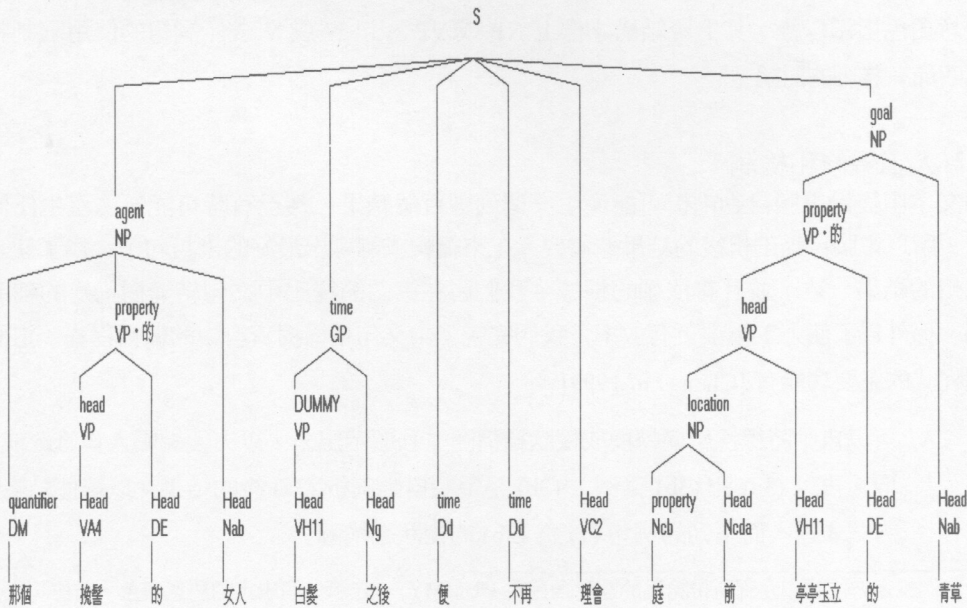
- B. 可以在樹狀結構圖中直接選取需要修改、移動或變更小結構樹 (sub-trees)，這個功能使編輯人員更有效率的產生正確的結構樹。
- C. 可以將電腦剖析程式無法產生的結果，藉由人工校正以後，加入句結構樹資料庫中，使我們剖析系統得以學習原本無法剖析的句型。
- D. 檢錯功能：目前此編輯程式尚未包含這項功能，但預計未來這些功能完備之後，不但可以檢錯，還可以自動修復不完整的句子以完成剖析或修正錯誤的剖析結果句。

(3a) 那個挽髻的女人白髮之後便不再理會庭前亭亭玉立的青草

(3b) 線性結構：

#S(agent:NP (quantifier:DM: 那個 | property:VP · 的 (head:VP (Head:VA4: 挽髻) | Head:DE: 的) | Head:Nab: 女人) | time:GP (DUMMY:VP (Head:VH11: 白髮) | Head:Ng: 之後) | time:Dd: 便 | time:Dd: 不再 | Head:VC2: 理會 | goal:NP(property:VP · 的 (head:VP (location:NP (property:Ncb: 庭 | Head:Ncda: 前) | Head:VH11: 亭亭玉立) | Head:DE: 的) | Head:Nab: 青草)) #

(7) 樹狀結構圖：



#### 4. 中文句結構樹建構輔助原則

在處理中文句結構樹時，常會遇到歧義結構的問題。在其他專家學者尚未提出具體有力的證據前，爲了有效解決這種狀況，並求有個統一的處理方法，我們提出三項輔助原則，分別以對等連接詞或並列結構、有多重修飾語的結構等爲例加以說明。

##### 4.1 詞類小而美

「詞類小而美」原則是爲了解決如何標示語義等重，但組成成分的類型不盡相同的詞組之語法類型，如對等連接詞詞組或並列結構。以下分三種情形來討論標記的原則：

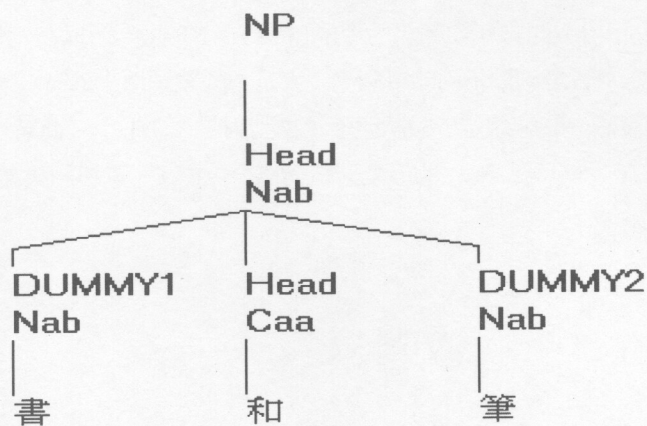
A. 下層結構皆由單詞構成，又可細分爲下列情況：

a. 詞類完全相同：

以 (8a) 爲例：「書」和「筆」皆是單詞而且詞類同爲「Nab」，也就是個體名詞，此時「書和筆」的上層結構標示爲「Head: Nab」，而「Head: Nab」的上層結構是「NP」，剖析結果如 (8b)。

(8a) 書和筆

(8b)



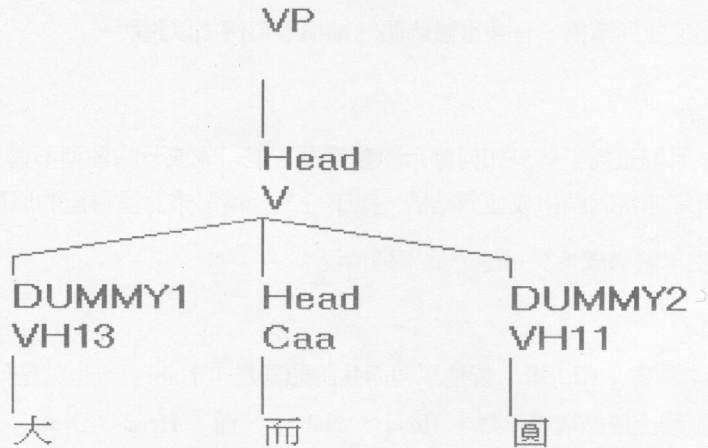
b. 詞類不完全相同：

在例句 (9a) 中，「大」和「圓」皆是單詞但是詞類不同，「大」的詞類爲「VH13」，也就是帶對象 (goal) 和範圍 (range) 的狀態不及物動詞；「圓」的詞類爲「VH11」，也就是一般的狀態不及物動詞。在這種主要類別相同，但是細類不同的情況下，中心語的語法形態只標上其主要類別；因此「大而圓」的上層結構爲「Head: V」，而「Head: V」的上層結構是「VP」，剖析結果如 (9b)。



(9a) 大而圓

(9b)

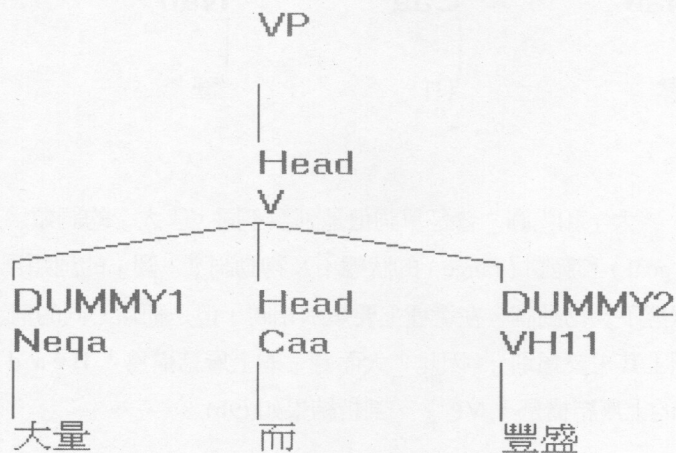


c. 詞類完全不同：

當下層結構由兩個不同詞類的單詞所構成時，我們規定上層詞組形式依 V、N、P、Ng 的順序取詞類；如下層結構由「V」和「N」構成時，上層結構為「V」；若下層結構由「N」和「Ng」所構成，則上層結構為「NP」。所以在例句 (10a) 中，「大量」和「豐盛」皆是詞類不同的單詞。「大量」的詞類為「Neqa」，是表示物品數量的定詞；「豐盛」的詞類是一般狀態不及物動詞「VH11」。此時「大量而豐盛」的上層結構為「V」，而「V」的最上層結構為「VP」，剖析結果如 (10b)。

(10a) 大量而豐盛

(10b)

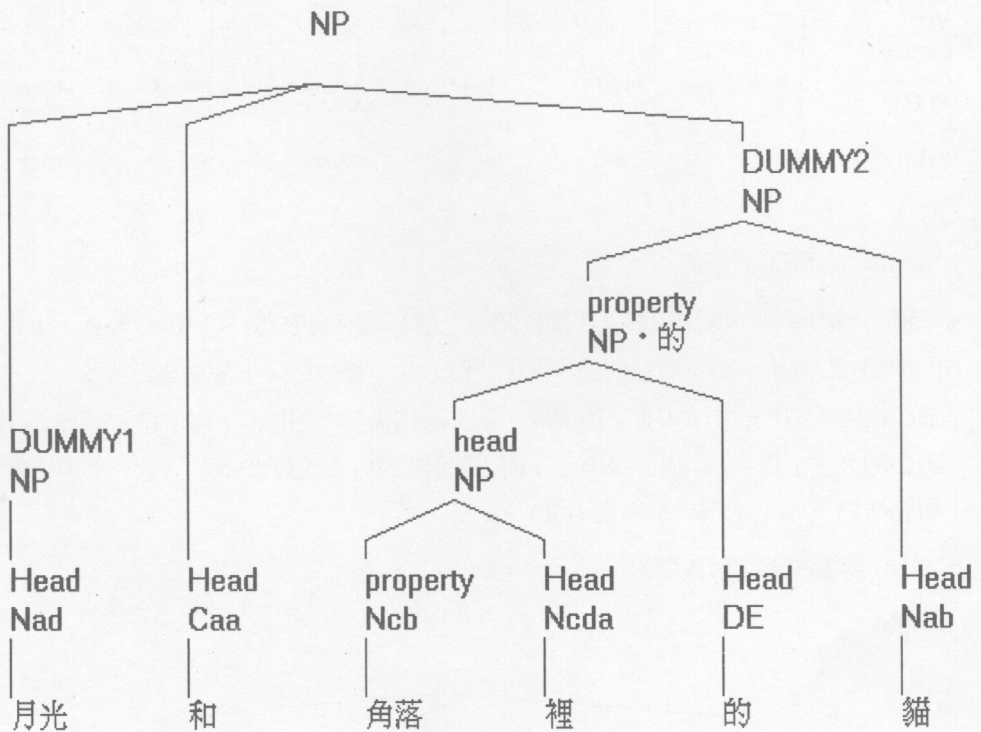


B. 最下層結構有一方是詞組，則取其詞組類別。

在例句 (11a) 中，「月光」是單詞，詞類是抽象名詞「Nad」；「角落裡的貓」是名詞詞組「NP」。此時，「月光」的上層結構為「NP」以呼應「角落裡的貓」，而「月光和角落裡的貓」的上層結構為「NP」，剖析結果為 (11b)。

(11a) 月光和角落裡的貓

(11b)



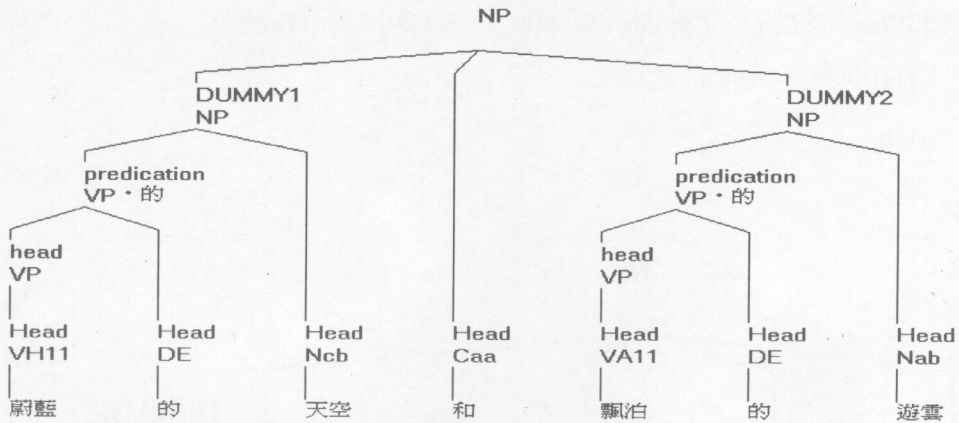
C. 最下層結構皆由詞組構成，則取詞組類別。

a. 下層結構的詞組形式相同：

例句 (12a) 「蔚藍的天空和飄泊的遊雲」中，由於「蔚藍的天空」和「飄泊的遊雲」皆為名詞詞組「NP」，所以它們的上層結構也是「NP」，剖析結果如 (12b)。

(12a) 蔚藍的天空和飄泊的遊雲

(12b)

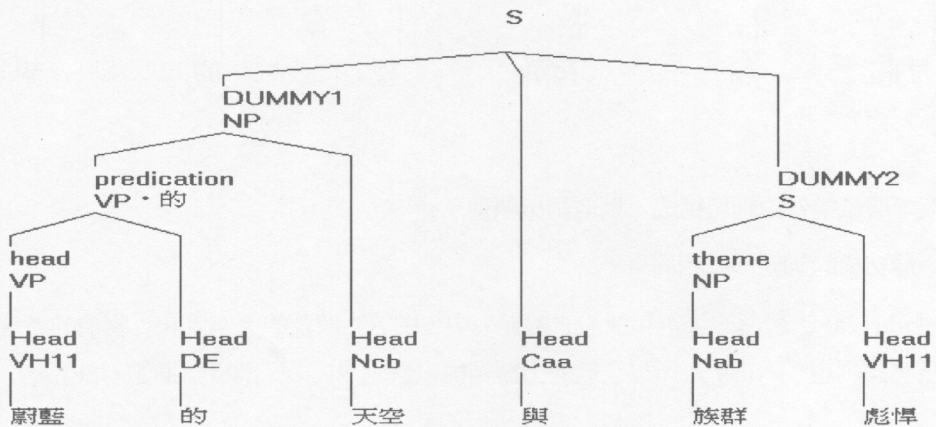


b. 下層結構的詞組形式不同：

當下層結構由兩個不同詞組形式所構成時，上層詞組形式依 S、VP、NP、PP、GP 的順序取詞組。如下層結構由「VP」和「S」構成時，上層結構為「S」；若下層結構由「NP」和「GP」所構成，則上層結構為「NP」。以 (13a) 為例，由於「蔚藍的天空」是名詞詞組「NP」，而「族群彪悍」是句子形式「S」，所以它們的上層結構為「S」，剖析結果為 (13b)。

(13a) 蔚藍的天空與族群彪悍

(13b)



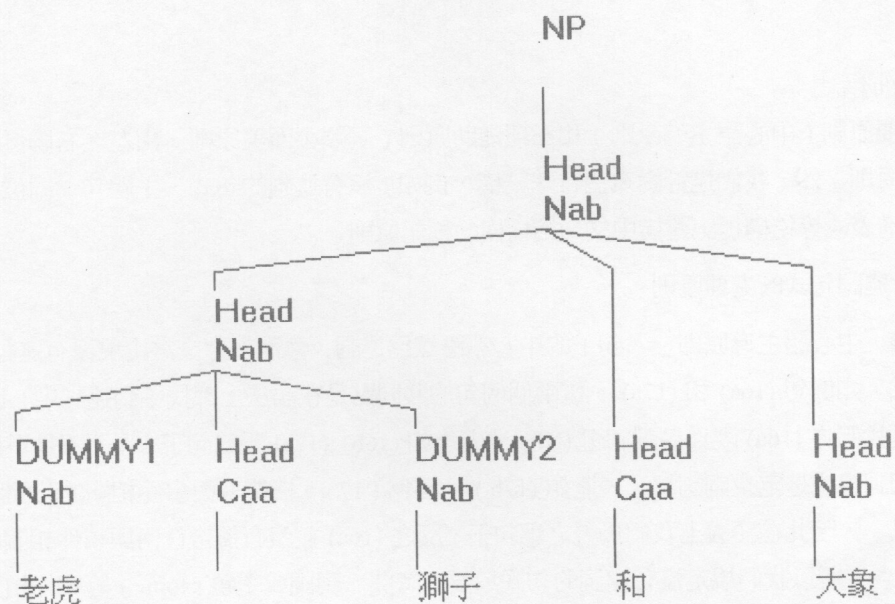


#### 4.2 由左至右聯併原則

在「對等連接詞詞組」或「並列結構」裡，組成成分的語義等重時，由左至右一一聯併。以(14a)為例，先將「老虎、獅子」視為一個詞組，分別與「大象」這個詞組同時扮演中心語「和」的必要論元，然後再以「、」為中心語，「老虎」與「獅子」各為中心語「、」的兩個必要論元。所以就層次而言，最右邊的詞組將與左邊的詞組構成上層結構，而後由左邊詞組的最右邊詞組及左邊詞組構成第二層結構。所謂「由左至右聯併」正是架構「對等連接詞詞組」或「並列結構詞組」的原則。除了由左至右聯併原則外，(14a)還依據詞類小而美原則來標上其語法形式，剖析的結果如(14b)。

(14a) 老虎、獅子和大象

(14b)

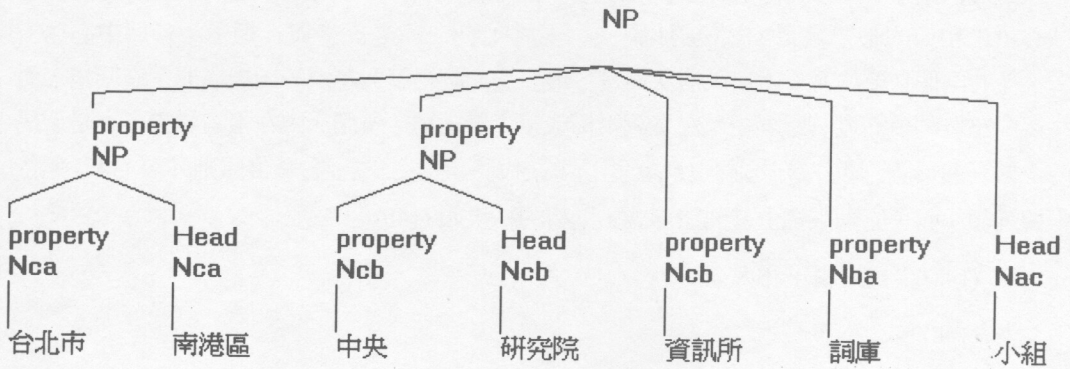


#### 4.3 扁平原則

根據「中心語主導原則」，每個詞組結構皆由中心語與其論元或附加成分組成。而「扁平原則」則使得每個詞組結構中的論元或附加成分的最上層結構都與中心語位於同一層次。由於中文語義的重心大多位於最右邊，故剖析系統會由最右邊的詞組聯併到左邊，而在中心語左方的詞組則以語義為單位來組成中心語的論元或附加成分。以名詞詞組為例，(語義上的)中心語大多位於最右邊的單詞(詞義獨立而完整的詞)，而左邊的單詞則構成中心語的修飾成分。在例句(15a)中，「小組」是中心語，而「台北市」、「南港區」、「中央研究院」、「資訊所」、「詞庫」皆是中心語的修飾成分。由於「中央」是「研究院」的修飾語，不屬於「小組」的修飾語，故將「中央研究院」，和「資訊所」、「詞庫」、「小組」置於同一層次，共同修飾中心語「小組」。

(15a) 台北市南港區中央研究院資訊所詞庫小組

(15b)



#### 4.4 其他結構

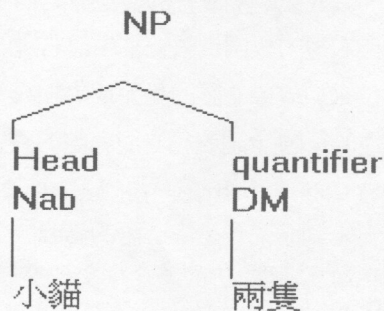
除了主要原則（中心語主導原則）和三項輔助原則（詞類小而美原則、由左至右聯併原則和扁平原則）外，我們還針對幾個特殊結構，例如：沒有述詞的句式、主題句、同位語和整體結構表義等結構提出建構中文句結構樹的標記原則。

##### A. 沒有述詞句式的處理原則

依據「中心語主導原則」，句子的中心語應該為述詞，然而在中文常出現沒有述詞的句式，如例句 (16a) 和 (17a)。這兩個例句的共同點是在語法上都找不到述詞中心語。但是我們將 (16a) 視為名詞詞組 (NP)，表達如 (16b)；(17a) 視為句子 (S) 而不是單純的名詞詞組或是定量詞詞組，表達如 (17b)，因為 (17a) 攜帶著與事件相關的時間狀語「已」，因此在語義上我們仍視之為句子；反之 (16a) 並沒有攜帶著任何與事件相關的語意概念。所以我們規定沒有述詞的句式只要有時間、副詞或主題 (topic) 等事件相關語意概念時，都視為句子 (S)；否則為名詞詞組 (NP)。

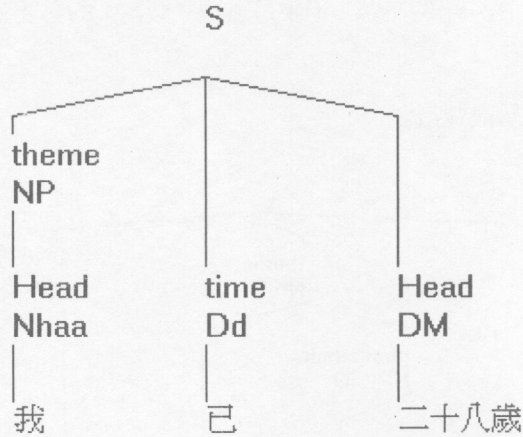
(16a) 小貓兩隻

(16b)



(17a) 我已 28 歲

(17b)



## B. 主題句的處理原則

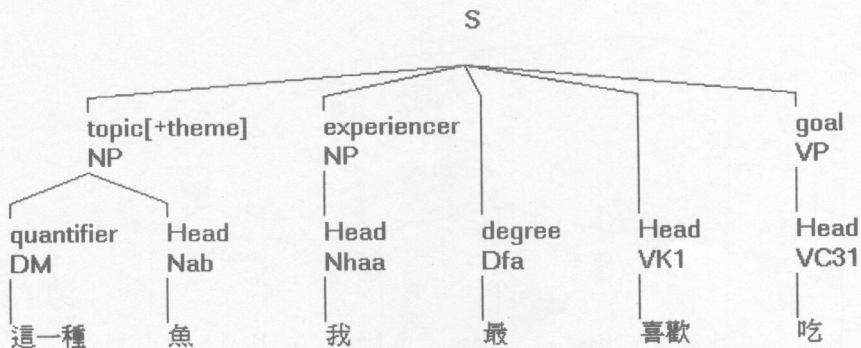
主題句 (topic sentence) 是中文常見的句式，但是其「主題」 (topic) 的語意角色有時並不容易決定，因為在句中，它與其他詞項之間的語意關係可能具有多重的角色，可能既是句子的「主題」也是述詞的必要論元或者只是單純扮演句子的「主題」而已。因此根據下列二種狀況，我們給予不同的標記原則：

## a. 同時是句子的「主題」和述詞的必要論元：

除了指派語意角色 *topic* 外，我們以附加特徵的方式來表達，標記為 *topic [+thematic role]*，如例句 (18a) 剖析為 (18b)，「這一種魚」是句子的主題也是述詞「吃」的必要論元，論旨角色為 *theme*，因此標上 *topic[+theme]*。

(18a) 這一種魚我最喜歡吃

(18b)

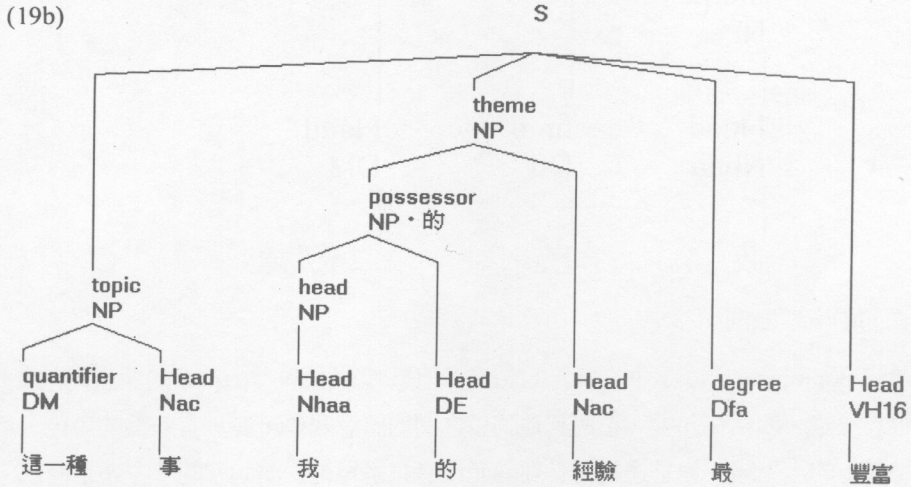




b. 是句子的「主題」但不是述詞的必要論元：

如果只是單純扮演句子的「主題」而不是述詞的必要論元，我們只標上 topic，如例句 (19a) 中「這一種事」是句子的主題，但不是述詞「豐富」的必要論元，因此標上 topic，結果如 (19b) 所示。

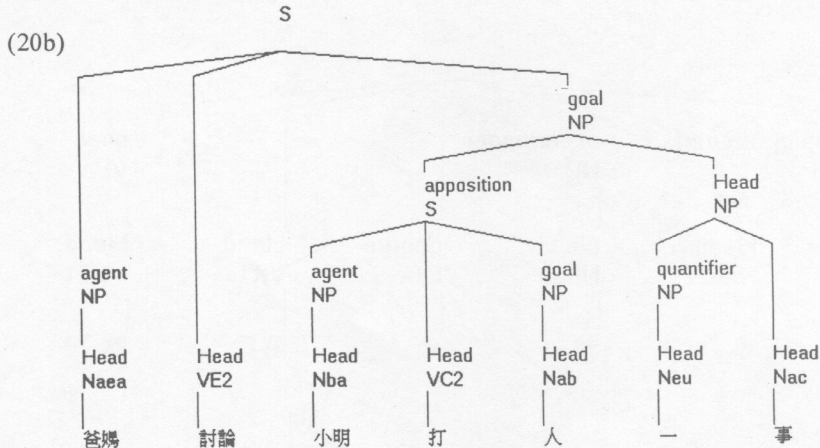
(19a) 這一種事我的經驗最豐富



C. 同位語的處理原則

同位語 ( apposition ) 是名詞詞組的語意角色，指涉與中心語同等義的詞組，與名詞詞組中心語具備相同的詞組形式，然而在中文裡常出現述詞詞組 ( 或句子形式 ) 成爲名詞中心語的同位語。由於同位語與名詞中心語等義，因此與名詞中心語構成雙岔結構，其詞組形式可以是句子或述詞詞組，而不拘泥於名詞詞組，如例句 (20a) 中，句子形式的「小明打人」是名詞詞組「一事」的同位語，剖析結果爲 (20b) 。

(20a) 爸媽討論小明打人一事

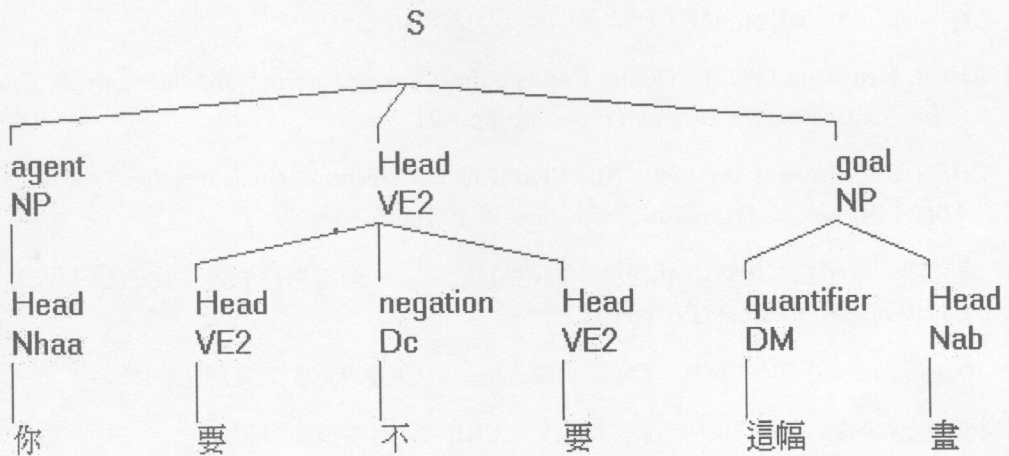


## D. 整體結構表義的處理原則

中文有時會遇到一些無法用單詞意義解釋，而用一個結構來表示其意義的情形，因為它所代表的是整個結構的意義，所以就必須以整體結構來表義。我們的處理方式是先標記出整個結構的意義，然後再分析其內部結構。如例句 (21a) 中，「要不要」是正反問句 (A-not-A 問句) 的結構，並且在這個句子中扮演主要動詞的角色，所以整個「要不要」為中心語，其語法類別為動作句賓動詞 (VE2)，之後再分析下「要不要」的內部結構，結果為 (21b)。再如 (22a)，「鞠躬二個躬」是內部結構為動賓結構的類單賓動詞 (VB) 「鞠躬」中插了「二個」，因此我們給予「鞠躬二個躬」(VB11) 的詞類標記，再分析其內部結構，並標上特徵 [+spv] 和 [+spo] 來表示「鞠」和「躬」分屬動賓結構動詞和賓語的部分，而之間在形式上雖然中插了其他成分，其整體意義仍然非常緊密，結果為 (22b)。

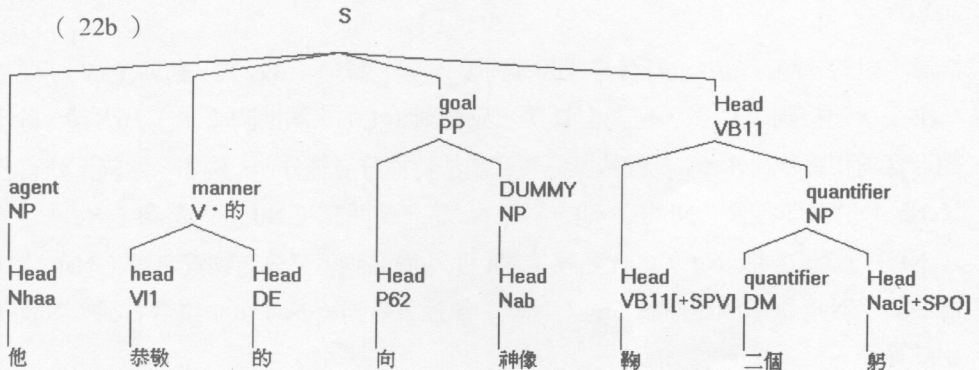
( 21a ) 你要不要這幅畫。

( 21b )



( 22a ) 他恭敬的向神像鞠躬二個躬。

( 22b )



## 5. 結論

以訊息為本格位語法表達模式的中文剖析系統，將每一個句子剖析成結構樹，並且表達出詞和詞之間複雜的語法和語意關係，再利用編輯程式，並且根據詞類小而美、由左至右聯併、扁平等編修原則，加以人工修正和檢驗，一步一步架構起中文句結構樹資料庫。我們從完成的 38,725 棵中文句結構樹中，開放一千棵結構樹以供學者專家在中文句法、語意關係研究參考之用。網址為 <http://godel.iis.sinica.edu.tw/CKIP/trees1000.htm>。

## 參考文獻

1. CHEN, Keh-Jiann. (1996) "A Model for Robust Chinese Parser." *Computational Linguistics and Chinese Language Processing*. vol. 1, no.1. pp.183-204.
2. CHEN, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, Hui-Li Hsu. (1996). "Sinica Corpus: Design Methodology for Balanced Corpra." *Proceedings of the 11th Pacific Asia Conference on Language, Information, and Computation (PACLIC II)*, Seoul Korea, pp.167-176.
3. CHEN, Keh-Jiann, Chu-Ren Huang (1994). "Features Constraints in Chinese Language Parsing." *Proceedings of ICCPOL '94*, pp. 223-228.
4. CHEN, Keh-Jiann (1992). "Design Concepts for Chinese Parsers." *3rd International Conference on Chinese Information Processing*, pp.1-22.
5. CHEN, Keh-Jiann, et al (1999). "The CKIP Chinese Treebank: Guidelines for Annotation." *ATALA Workshop - Treebanks*, Paris, June 18-19, pp85-96.
6. 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 ". *Proceedings of ROCLING II*, pp97-119.
7. 中文詞知識庫小組 (1993). 《中文詞類分析》. CKIP-93-05 中文詞知識庫.
8. 林甫雯 (1992). 《 ICG 中的論旨角色 》. CKIP-92-01 中文詞知識庫.

## 附 錄

### 1 詞類表

詞庫小組的分類系統中，訂有非謂形容詞 ( A )、體詞 ( N )、動詞 ( V )、介詞 ( P )、連接詞 ( C )、副詞 ( D )、感歎詞 ( I )、語助詞 ( T ) 八大類，除了非謂形容詞和感歎詞外，其餘依其語意和語法特性分層細分。例如，體詞下分有名詞 ( Na )、專有名稱 ( Nb )、地方詞 ( Nc )、時間詞 ( Nd )、定詞 ( Ne )、量詞 ( Nf )、方位詞 ( Ng )、代名詞 ( Nh )。而名詞下又分有物質名詞 ( Naa )、個體名詞 ( Nab )；個體抽象名詞 ( Nac )、抽象名詞 ( Nad )；集合名詞 ( Nae ) 五小類等等。



## 2 語意角色

語意角色包括中心語、論元及附加成分。因中心語不同，相關的論元角色及附加成分亦有不同。

A. 介詞和連接詞的必要論元是 DUMMY 。

B. 修飾述詞的常見語意角色：

aspect, time, location, benefactor, companion, comparison, complement, condition, conjunction, deontics, duration, epistemics, evaluation, exclusion, frequency, imperative, inclusion, instrument, interjection, manner, negation, particle, quantifier, quantity, reason, standard, target, topic, agent, theme, goal, causer, source

C. 名詞分爲二類：一般性名詞和名物化名詞

a. 修飾一般性名詞的語意角色

predication, property, apposition, possessor, quantity, quantifier

b. 修飾名物化名詞的語意角色

time, location, property, predication, quantifier, nominal, agent, theme, experiencer, goal, negation

D. 述詞論旨角色

theme, agent, causer, instrument, source, goal, range, location

## 3 論旨角色圖

論旨角色

