

訊息為本的格位語法一

一個適用於表達中文的語法模式

陳克健

中央研究院資訊科學研究所

黃居仁

中央研究院歷史語言研究所

摘 要

本論文提出一個以訊息為本的語法模式，這個語法模式和其配合的剖析方法，可以很精確的表達及很有效率的分析中文句子。本語法模式採用詞彙為中心的表達方式，將每一個詞的語法及語意訊息以特徵結構表示。詞彙結合為片語，片詞建構成句子，以中心語驅動，結合時必須符合句子中詞彙所規定的語法限制。語意的合成假設具有結合性可以從詞彙的語意訊息聯併獲得。此一語法模式保存了聯併語法的所有優點而且兼顧了剖析的效率與語意的分析。

一. 緒論

本論文的主要目的，在提出一個適合於表達中文的語法模式及其剖析方法。促使我們有這種想法的理由，是因為目前似乎沒有一個語法表達模式和與其配合的電腦剖析方法，可以很有效率的分析中文的句子。我們心目中理想的語法模式，在表達上應可以達到十分精確的地步，可以避免一般剖析時語法蔓生的現象 (*Over-generate*)。這個語法模式所表達的語法，不但可以用來分析語言，同時可以產生語言。而利用此一語法分析的結果，可以是一個良好的語意的表達形式。從電腦處理的觀點而言，這個語法模式的剖析程式，必須是簡單易行，而且不因語法規律的改變而改變剖析程式。也就是說，它具有宣稱性 (*Declarative*)。從認知的觀點而言，這一個模式能模擬人類語言認知的現象。*Chomsky* 認為人類與生就有一個處理語言的器官 (*Organ*)。這點暗示我們，電腦的剖析器就是相對應的語言器官。很顯然地，這個器官不能經常改變

，結構也不能太複雜，否則電腦系統的設計就會大有問題。學習語言，不是在學習如何改變這個器官，應該在於學習不同詞彙的用法；如何將詞彙組合成句子；如何從詞彙知識分析句子的結構，進而了解句子的意義。日積月累的結果，我們學習了更多的詞彙及詞彙的相關語意及語法訊息，使我們的語言能力日益精進。同樣的道理，對電腦而言，如何輸入、整理、儲存，甚而學習、推演 (*Inference*) 每個詞項的語意語法訊息，將是語言系統增強語言能力的主要方式。詞彙就像是一個個不同形式的積木，我們利用這一個個的積木來建造合法的句子，建造或分析的方法可以十分簡單。這個觀念可行的理由，是建立在詞彙結合成句子時具有意義上的結合性這假設上。近代語法理論中，以聯併 (*Unification*) 為基礎的處理方法 [Shieber 86] 和這個簡單構想十分相似。

我們可以假設，詞彙建構成片語及句子的規律訊息都附屬在每一個詞上，如 (1) 所示：

$$(1) \quad \text{詞:} \quad \begin{bmatrix} \text{語意訊息} \\ \text{語法訊息} \end{bmatrix}$$

每一個詞的附屬訊息，描述了這個詞如何扮演句子的構建成分。聯併的作用，即在確定詞串中的每一個詞都依據其語法及語意訊息，以正確的順序及形式，扮演其所應扮演的成分角色。本篇論文將從訊息的表達方式，分析的方法及各個不同理論的比較，來逐步說明此一適用於中文剖析的語法理論，進而研究並希望此一語法理論，可以適用到任何的語言上。

這個語法的產生擷取了許多語法理論的特點，很難追根究底的探究某一個構想來自何方。不過從醞釀到完成，影響我們最大的語法理論有 *Case Grammar* [Fillmore 68]，*Lexical Functional Grammar* (LFG) [Bresnan 82, Sells 85]，*Generalized Phrase Structure Grammar* (GPSG) [Gazdar 85, Sells 85]，*Functional Unification Grammar* (FUG) [Kay 85]，*Categorial Unification Grammar* [Uszkoreit

86] , *Head-driven Phrase Structure Grammar* (HPSG) [Pollard 87] 。特別是 HPSG, 我們的語法理論上可以看到許多 HPSG 的影子。有些是我們自己的想法和 HPSG 不謀而合, 有些靈感是來自 HPSG 。

二. 語法語意訊息的表達

我們先從句子的表達方式來探討那些語法語意的特性足以描述一個合法合理的句子。一個句子的中心語是這個句子的主要動詞, 複雜句可以看成是由幾個句子結合而成的。因此在語意上而言句子構成了命題 (*Proposition*) , 每一個命題由中心動詞—述語 (*Predicate*) 及其論元 (*Arguments*) 所組成。在語法上而言, 句子是由動詞 (*Verbs*) 及其論元及附加成分 (*Adjuncts*) 依據一定的順序排列而成。因此我們採用了 Fillmore 的 *Case Grammar* 的架構做為句子的語意表示法。如 (2)

$$(2) \quad \begin{array}{l} \text{命題:} \\ (\text{Proposition}) \end{array} \left[\begin{array}{l} \text{述語 (Predicate)} \\ \text{論元 (Arguments)} \\ \text{附加成分 (Adjuncts)} \end{array} \right]$$

述語為句子的主要動詞; 論元及附加成分以其語意格位表之, 語意格位的指定, 因受動詞支配且和名詞本身及常識皆有關連, 因此角色精確程度可依實際狀況調適。句子經過剖析以後, 希望能得到 (3) 的結果, 且 (3) 的表示形式經過產生的過程能得到對應的句子。

(3) 張三昨天在公園打了李四

$$\text{命題:} \left[\begin{array}{l} \text{述語: 打} \\ \text{時間: 昨天} \\ \text{地點: 公園} \\ \text{主事者: 張三} \\ \text{對象: 李四} \end{array} \right]$$

上一節裡, 我們提出用聯併的方式, 從詞彙的語法語意訊息得到 (3) 的結果。聯併可以看成是一種訊息的累積及檢測的過程。我們

假設詞彙構成片語，片語構成句子時都有意義上的合成性 (*Compositionality*)，因此可以適用訊息累積的方式。語法上的詞彙結構關係，位置關係，共存限制，形式限制，都可以視為一些限制條件 (*Constraints*)。聯併過程就是語法訊息的檢測，因此接下來必須分析構成句子的這些訊息是如何從詞彙上得到，詞彙上應如何安排各個詞彙語意語法訊息。

在討論詞彙訊息以前，我們先對一些語法上的限制條件加以說明。

2.1 語法的限制條件及表達方式

傳統的語法表達方式多採用無語境的詞組結構律 (*Context-free Phrase Structure Rules*) 為主，因為自然語言除了少數的特殊現象外，都可以歸類為無語境語言。計算語言學家為了表示上的簡便性，也發展出許多強化無語境語法 (*Augmented Context-free Grammars*)，如 *ATN* [Woods 70, Winograd 83]、*GPSG* [Gazdar 85]、*LFG* [Bresnan 82] 等。

如果一個語言可以單純的以無語境詞組結構律精確的表示出來，則每個詞或成分的語法訊息，就是有關於這個詞或成分的語法規律。但是這樣的表示法，具有以下缺點：

1. 只能表示分類的共同訊息，個別差異無法修正。若有個別差異必須自成一類。以動詞為例，動詞片語的規律必須因及物性的不同分為：不及物、及物、雙及物，及物動詞又因賓語不同分為及物接名詞、及物接動詞片語、及物接句子等等，而產生許多條相似規律，但這些規律又不能滿足各個動詞的個別差異，是不是要再分下去，再把規律增加呢？

2. 一致性及共存限制等問題也必須以相似規律重複表示。

例如英文裡為了表示主語和動詞的一致性，必須有兩條規律

$$S \rightarrow NP1 \quad VP1$$
$$S \rightarrow NP2 \quad VP2$$

NP1 代表第三人稱單數名詞片語，NP2 代表非第三人稱單數名詞片語，分別後接第三人稱單數動詞片語 VP1，及非第三人稱單數動詞片語 VP2。中文裡名詞和量詞之間也有一致性的問題。

3. 無法兼顧語意的訊息。

強化無語境語法可以採用一些語意的特徵，但還是把語法和語意分開處理。

4. 表示法不夠精簡，對於詞序比較自由的語言，表達時更為繁瑣。

例如：A 是由 B, C, D 組成，但 B, C, D 詞序不受限制。用無語境語法必須排出所有組合：

$A \rightarrow BCD \quad A \rightarrow CDB$

$A \rightarrow BDC \quad A \rightarrow DBC$

$A \rightarrow CBD \quad A \rightarrow DCB$

但是用直接支配律 [Gazdar 85] 只要一條規律 $A \rightarrow B, C, D$ 。

我們摒棄了傳統詞組結構律的表示方法，採用了近代語法理論的一些優點，例如以直接支配律 (ID rules) 及線性律 (LP rules) 取代無語境語法規律 [Gazdar 85]，以特徵集合代表詞類 (Categories) [Gazdar 87]，以語法限制性 (Constraints) 取代衍生律。亦即若無限制條件則表示許可，來代替一般語法能用規律產生者才是合法的方式。這些語法的表達細節將在下一節詳細說明。

2.2 詞彙的訊息內容

每一個詞項下有其語法及語意的訊息，我們可以用特徵集合表之：

(4)	語意 訊息	意義：
		特徵：
		論元：
		附加成分：
	語法 訊息	詞類：
		語法限制：
		成分形式 (Form)
		基本句型 (Basic Pattern) :
		成分規律 (Adjunct Precedence) :

動詞的特徵集合，事實上代表了以這個動詞為中心語的所有簡單句子所組成的集合。名詞的特徵集合代表了以這個名詞為中心語的所有名詞片語所成的集合，同理，介詞、方位詞、定詞，也都具有相同的特徵結構只是各個不同的詞類為中心語所表達的片語結構不同，因此特徵值 (feature values) 會有不同，非片語中心語之詞類亦有相似特徵結構，只是欠缺語意論元及附加成分，因此也沒有語法限制的訊息。我們以動詞‘勸’為例 (5) *：

(5) 勸	
語意 訊息	意義：‘勸’
	特徵：言談類
	論元：
	主事者：{ 意義：？ 特徵：人類 對象 []：{ 意義：？ 特徵：人類 題旨：{ 意義：？ 特徵：動作動詞 論元：主事者：[] 附加成分：時間、地點、法相、副詞修飾語、語尾助詞、 其它具有角色標記 (Case marker) 的介詞片語
語法 訊息	分類：Vnp, vp /* 雙及物動詞且直接賓語為 VP */
	語法限制：線性律：{ 主事者 [NP] < * < 對象 [NP] < 題旨 [{NP, VP}] 時間 < 介詞片語 < * < 語尾助詞

* 為了解釋上的簡便，例 (5) 中“勸”的語法訊息及語意訊息僅列出最主要的一小部分。

這個特徵集合和無語境語法比較，可以發現 *ID*, *LP* 的訊息隱含在裡頭，動詞及其論元構成了 *ID* 成分，*LP* 規律及語法形式表示在線性律上。在線性律上，每一個角色並可附帶其它形式上或語意上的訊息。例如

(6) *AGENT* [*NP*, *PLURAL*] 是 $AGENT : \left[\begin{array}{l} CAT : NP \\ NUMBER : PLURAL \end{array} \right]$ 的簡寫。

至於一些非詞彙類別 (*Lexical categories*) 如 *NP*, *VP*, *S*, *PP* 等定義為以中心語加上其它成分，並滿足每一成分中的語法限制的詞串 (*Word string*)。如何滿足語法限制的細節部分將在 2.4 節中討論。因此句子中心語為主要動詞，在主要動詞的語意及語法訊息中規定了句子的成分及結構。而成分的結構則由成分的中心語規定之。

此一語法理論的第二個特色，是角色名稱採用語意角色為主。基本上，角色名稱是以訊息為基礎的表示法，給予的訊息可以是語意角色，可以是語法角色，也可以附加語法特徵 (*Syntactic features*) 或語意特徵 (*Semantic features*) 如 (6)。訊息為本的描述方式更能精確的掌握每一個成分角色的語法及語意訊息。這個特點正是表達中文語法的一個必要條件，例如中文的時間成分語法結構非常複雜，有名詞片語 (*NP*) 形式，如 '早上'，有定量複合詞 (*DM*)，如 '上星期'，有方位詞片語 (*GP*)，如 '開會後'，有介詞片語 (*PP*)，如 '直到今天'，有副詞 (*Adv*)，如 '曾經'。但有共同的語意特徵 '+time'。如何把時間成分描述得清楚呢？以訊息為主的表示法較具有彈性，然而必須是電腦可以了解的形式。我們採用系統網路 (*System network*) 的方式 [Winograd 83]，在系統網路上選擇訊息特徵，使得每個角色成分為一組可解析的 (*Interpretable*) 特徵集合。系統網路的細節在 2.3 節說明。時間成分依據系統網路可描述如下：

(7) *time*[[*NP* [*Nd*],

DM[*Nfge*, +*definite*],

GP[{前, 以前, 之前, 後, 以後, 之後, 內, 以內, 之內, 裡,

左右, 中, 之中, 間, 之間, 中間, 當中, 底, 頭,

末, 終, 初, 起, 來, 以來, 開始, 時, 之時, 之際}],

PP[{在, 於, 到, 直到, 等到, 趁, 臨, 趕, 自, 從, 自

從, 由, 打, 打從, 當}],

ADV[*Dd*]]]

這一個語法理論的另一個特色是，語意成分和語法成分為詞彙的兩個主要訊息，語意成分的訊息是放諸四海皆準的，不管任何語言，詞彙都具有其語義及其意義類別，而支配的論元是完成一個完整訊息所必需的其它語意成分。以‘勸’為例，完成‘勸’這個動作必須有一個主事者，一個對象及一個題旨，其它論元代表了可能修飾的成分。這些成分角色也和意義有關，不因其為何種語言而有所改變。詞彙組合成片語、句子以後，片語和句子的組合語意訊息應該能反應片語或句子的意義。我們定義語意上的必要論元，為構成一個完整的訊息在觀念上必然存在的成分，例如動詞“打”有兩個必要論元，一個是主事者，另一個是打的對象（GOAL），但是語法上一個合法的句子並不一定要存在所有的必要論元，例如“我打了”，打的對象可以不出現在句子上，但是觀念上這個對象還是存在的。所以在我們的語法模式中我們將語意上的必要論元和語法上的必要論元分開。語法上的必要成分和語言不同有關，我們以基本型（*Basic Pattern*）表示出可能的中心語和論元組合，2.4 節會有較詳細的說明。語意的成分可維持普遍性（*Universal*），同一個觀念（*Concept*）（相對於一個詞）任何語言有相同的表示結果。但是非句子層次的必要論元通常不一定，是因為要完成完整的意義，可能是語法上的需要，例如定詞‘這’必須有量詞這個論元，所以‘這’的必要論元是屬於語法訊息，而非語意訊息。因此也是用語法欄內的基本型表示。

意義一欄內表示的是詞彙的意義，在尚未找到適合的表達語言 (Representational Language) 的情況下，我們認為中文是一個相當良好的中間語言，因為中文詞彙的意義相當確切，不像英文有多義的情況。

語法訊息表達的是個別語言的特性，不同的語言有不同的語法特性，因此我們將它和語意訊息分開。它包含有詞類及語法限制 (Constraints)。語法限制包含了形式上的限制及詞序上的限制。詞序的限制我們一概以線性律表示。

2.3 語意角色的描述方式及其指派關係

直接採用語意角色作為線性規律角色成分名稱的理由如下：

1. 中文語法角色主語 (Subject)，賓語 (Object) 不明確。如例 (8) 無法清楚的決定主語、賓語。

(8) a. 這本書 李四看了。

b. 門口來了兩個人。

c. 箱子上綁了一條繩子。

d. 昨天下雨了。

2. 成分之間的順序關係和意義較有關聯，和語法結構關係較小。例如時間成分，如上節所述，可以是 (1) 名詞 (2) 定量式複合詞 (3) 介詞片語 (4) 方位詞片語 (5) 時間副詞。但是不管這些時間成分是什麼樣的語法形式，他們出現在句子中位置都是相同的。地點也有類似的情形。因此在語法表達上與其使用結構名稱還不如採用意義角色。
3. 最終的語意表達依然必須找出每一個成分的語意角色，而這些角色的指派又和語法結構有密切的關係，所以從電腦處理的着眼點而言，何不將語意角色的指派和語法結構分析結合在一起。

語意角色的指派和以下四種因素有關：

- a) 動詞的支配關係。
- b) 詞彙本身。
- c) 角色標記 (Marker)，如介詞‘被’是主事者標記。
- d) 詞序。

例如：

(9) 門口來了兩個人。

主事者角色‘兩個人’是由動詞‘來’，及‘人’的語意和位置關係決定。

目前大多數語法理論最弱的一環，就是在語意角色指定上。因為 LFG, GPSG 語意角色的指派，是留待語意分析時才處理。對英文而言，語法結構清楚後，語意角色可以從語法結構得到。對其它語言並不如此，例如日文。因此 Gunji 修改 GPSG，使角色指派取決於後置詞 (Postposition) 的標記，得到一種適用於日文的語法理論 JPSG [Gunji 86]。HPSG 的理論也注重語意的訊息，但是，對於語意角色的指派並無深入的探討，處理英文時還是以語法結構決定。

我們的語法理論將四種決定因素充分的表示在詞彙的語意及語法訊息內。以 (5) 為例，動詞‘勸’的語意訊息內表示出句子的必要論元及其語意限制，角色有主事者，對象及題旨，而附加成分欄內說明可能出現的選用論元。這些論元的語法類別，線性順序和標記都表示在線性律中，因此在剖析過程中，被指派的角色除了要符合語法規律外，語意的限制也被用來判定何者角色指派較為合理。

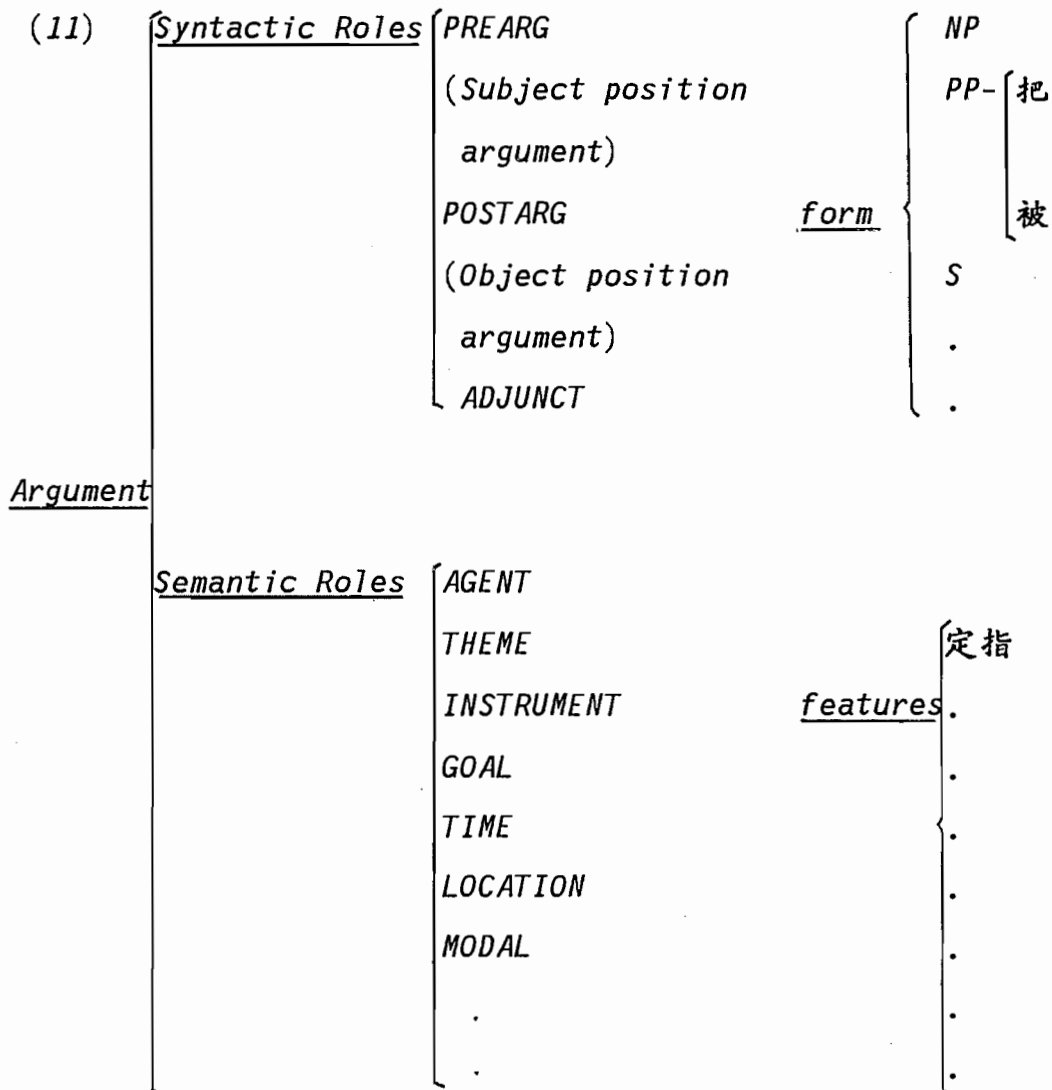
例如：動詞‘看’的語法訊息中可能有如下選擇性之線性規律。

(10) AGENT [NP] < ‘看’ < GOAL [NP]

{ AGENT [NP] , GOAL [NP, 定指] } < ‘看’ << COMP [ASP]

對例句‘這本書我看過’和‘我這本書看過’角色的指派必須依靠語意訊息及線性規律。詳細的中文語意角色指派關係可參考 [Chen 89]

角色的名稱結構，我們採用了系統網路 (11)，以選擇的方式訂定可能的角色名稱。所採用的符號也以 Halliday 的 *Systemic Grammar* 為主 [Winograd 83]。



2.4 句子的層次結構及線性規律

這個語法架構下的句子結構仍然保存了傳統的句子層次結構，只是在句子層次上不把句子分解為 NP VP 的組合，直接以動詞及其論元和其它修飾語形成一個層次。第二層以下是論元的層次結構及修飾

語的結構。因為 *ID*, *LP* 規律保持了傳統語法理論的特質，所以依然具有遞迴 (*Recursion*) 的性質，一個句子可能有很深的層次架構。在例 (5) 中‘勸’為主要動詞的句子，包含了動詞論元，主事者，對象，及題旨，及其它附加成分。題旨本身又是一個句子的架構，可以一直遞迴下去，同樣的，主事者等角色也有可能是關係子句，因此也可能有複雜的層次結構。這個層次結構看起來像是建立在語意架構上的，事實上和語法架構的差別在於語法層次上的成分有先後關係，成分的名稱為語法結構名稱如 *S*, *NP*, *VP* 等，因此在這個系統底下，我們也必須規定怎麼樣的詞序付以構成合法的成分 (*Constituents*) 如 *NP*, *VP*, *S* 等。

每一個詞彙項下有一個語法分類特徵。語法分類帶有兩個主要的訊息，其一是表示此類詞彙具有共同語法特徵，其二表示此一詞彙可能構成何種類型片語結構的成分角色。我們採用了「*X*標楨理論」(*X-bar Theory*)。每一個片語結構是由中心語 (*Head*) 及其修飾成分構成，片語結構的成分及其線性順序主要在片語的中心語上表示。對中文而言主要的片語結構有句子 (*S*)、介詞片語 (*PP*)、名詞片語 (*NP*)、形容詞片語 (*AP*)，方位詞片語 (*GP*) 和定量式複合詞 (*DM*) 等。由中心語的訊息可以知道在這個層次上如何構成合法的片語，句子的中心語是主要動詞，因此在動詞的訊息結構上顯示了句子的必要成分和可能的附加成分，同時用線性規律規範了成分之間的順序關係。介詞片語的中心語是介詞，因此在每一個介詞的詞彙項下具有以下訊息：有一個必要論元，一般是名詞片語，稱之為「被標記角色」(*Marked elements*)，有一個附加成分叫後置成分，一般是方位詞，如上、下、中、前、後等。具有的線性規律是 [介詞] < [被標記角色] < [後置成分]。這和規律 *PP* → *P NP (LOC)* 所表達的訊息相類似。只是我們的表示法可以更精確的表示出各個角色的語法結構及語意限制。名詞片語的中心語是名詞，每一個名詞加上一些附屬成分都可能形成名詞片語，每一個名詞上都帶有如何構成一個名詞片語的語法訊息。因此片語的存在由以下的中心語主導原則規範之。

中心語主導原則 (Head-driven Principle) :

句子或片語結構皆有一中心語 (Head) 。中心語的語意訊息欄內規定由此一中心語組成的成分結構，包含論元角色及附加成分，相當於一個直接支配律所表達的訊息。中心語的語法限制欄內規定，角色成分的形式限制，線性順序及共存限制。中心語的語法分類欄內規定此一中心語所組成片語結構的名稱。中心語若為動詞 (V) ，形成的片語結構為句子 (S) (包含動詞片語 $VP=S/Subject$) ；若為名詞 (N) 則形成名詞片語 (NP) ；若為介詞 (P) 則形成介詞片語 (PP) ；若為方位詞 (G) 則形成方位詞片語 (GP) ；若為形容詞 (A) 則形成形容詞片語 (AP) ；若為定詞 (D) 則形成定量式複合詞 ($DM=NP/NP$) 等。

線性規律規定了片語結構內每個角色成分必須遵守的順序關係。同一個角色可能有不同的順序位置。同一個角色可能有不同的語法結構，例如主事者角色可能是 NP 結構，可能是 PP [被] 結構，結構不同時，線性順序可能不同。在上一節中，我們已經說明了角色的特徵結構的系統網路，這裡我們只要提供線性順序的一些表示符號即可。

線性律使用的符號：

< : 例 $a < b$ ，角色 a 必須在角色 b 前出現。

<< : 例 $a << b$ ，角色 b 必須緊接在角色 a 後面。

{ } : 集合符號，例： $a < \{b, c\}$ ， a 必須在 b 及 c 之前， b 和 c 之間無順序限制。

{ : 選用符號，例： $\left\{ \begin{array}{l} R1 \\ R2 \\ R3 \end{array} \right.$ 滿足 $R1$ ， $R2$ ，或 $R3$ 中任一限制即可

\$: 存在符號，例 $a < * < \$ b$ ，在 b 存在的語境下， a 方能出現。

& : and 符號，例 $a \& R1$ ，在 $R1$ 的語境下， a 方能出現。

>< : 共存限制符號，例 $a >< b$ ， a 和 b 不能共同出現。

: 邊界符號，例 $\# < a$ ， a 必須出現在句首。

* : 中心語符號。

一個合法的片語結構必須滿足結構合法性原則，如下所示：

1. 必要論元存在而且唯一存在（相當於 LFG 之 *Completeness* 與 *Functional Biuniqueness*）。不過這裡所謂必要論元的存在與否，指的是論元結構只要符合某一基本句型即可，而不強求每一必要論元真正在句中出现。
2. 不得存在非容許之其它成分（相當於 LFG 之 *Coherence*）。
3. 每一成分必須符合成分間所有之線性規律。

2.5 句子的合理性與語意特徵的傳承

中心語的訊息結構中不但規定了什麼是合法的片語結構。同時利用語意訊息欄內論元的語意限制來規範片語結構的合理性。由於中文語句的語法結構具有相當程度的歧義 (*Ambiguity*)，可能分析為數個不同的語法結構，因此語意的訊息不但規範了結構的合理性，最主要的目的在於幫助解決語法歧義的問題。同時也提供剖析器選擇合理結構的參考。

片語結構的語意特徵是如何得到的呢？又是如何規範呢？基本上每一個片語應該繼承其中心語的語意特徵，例如‘美麗的女孩’應該具有‘女孩’的‘人’及‘女性’等特徵，另外有一些特徵的繼承不是從中心語來的，也有從論元或附加成分繼承的。例如‘誰來了？’是一個疑問句，疑問的特徵是從論元‘誰’承襲而來的。連接詞片語的特徵是從連接詞所連接的兩個論元所得到的。‘一本書’的‘-definite’及‘-plural’特徵則是承襲附加成分定量詞‘一本’的特徵。但是特徵的傳承並非僅限於簡單的繼承 (*Inheritance*)，有的時候是由推演 (*Inference*) 得到的。例如‘開會後’是一個方位詞片語，表示一個時間成分，如何從‘開會’及‘後’傳承 *+time* 的特徵呢？事實上‘開會’具有 *+EVENT* 特徵，‘後’具有 *+LOCATIVE* 特徵，在 [張88] 論文中規定 $[+EVENT] \& [+LOC] \supset [+Time]$ 。也就是說如果是 *EVENT* 和時空方位詞構成一個方位詞片語時，它們隱含時間特徵。這種推演出隱含特徵的情形在方位詞片語及介詞片語組成時最常見。這些複雜的特徵承傳應如何在語法模式中規定呢？我們有如下的特徵承傳原則規範之。

特徵承傳原則 (Feature Propagation Principle) :

- 1) 片語或句子的特徵應直接表達在中心語的語意特徵欄內。唯表達的特徵形式可能是 (a) 中心語特徵，以特徵值 (Feature value) 表之，如‘人’，(b) 論元或附加成分之特徵，以特徵路徑 (Feature path) 表之；如‘Argument feature’，‘DM feature’。
- 2) 另一類的特徵具有自動傳承至句子層次的特性，不必像其它特徵必須表明 (Explicitly specify) 在片語的中心語上。例如‘+Question’，‘+Negation’。自動傳承的特徵在傳承過程當中可能被阻擋 (Block)。被阻擋的情形有二，一種是直接阻擋，在特徵欄內標示阻擋的特徵‘IQ’。例如動詞‘研究’構成的句子，不具有疑問特徵，不管論元中的包接句是否帶有疑問特徵。因此就在‘研究’的題旨 (THEME) 角色的特徵欄內標示‘IQ’，可以阻擋可能自動傳承上來的‘+Question’特徵。另一種方式是吸收阻擋，通常‘+negation’的特徵只傳一層，由中心語所形成的片語所吸收，不再往上傳。
- 3) 推演形式的特徵承傳，採用分別列舉形式，以編號分別不同的情形，表達在中心語上。例如 (12)

(12) 前 :	Semantic :	{ feature: 1.+Time 2.+Location	
		argument: DUMMY	
	Syntax :	{ form: { 1.DUMMY[+Event]	
		2.DUMMY[+Physical]	

第一種情形‘前’的論元是‘事件’ (Event) 時，組成的方位詞片語具有時間 (Time) 的特徵，例如‘開會前’。第二種情形‘前’的論元是實體 (Physical) 時，組成的方位詞片語具有地方 (Location) 的特徵。

2.6 詞彙規律 (Lexical Rules)

詞彙規律的作用是将規則性的句型或片語結構變化，以規律的方式作用在線性律或論元上以得到表現新的句型或片語的成分律及線性律。例如：

1. 被動式：

[主事者] < [動作動詞] < [受事者] → [受事者] < [主事者，
pp [被]] < [動作動詞]

2. 把字句：

[主事者] < [處置動詞] < [受事者] → [主事者] < [受事者，
pp [把]] < [處置動詞]

3. 賓語省略： *Optional* (受事者)。

詞彙規律的另一個作用是分別包接句的容許句型和主要句子的不同。一般而言，主要句子結構較為靈活，包接句的句型較限定，特別是包接一個 VP 的句型，VP 是標準句型 *S/ SUBJ*，因此必須分別所謂標準句型及變化句型的適用範圍。採用詞彙規律是一種可能的分辨方式。

三. 系統設計的問題

3.1 詞彙訊息的表示方法

以上的語法理論，假設在每一個詞彙項下，包含了有關於這個詞的所有語法及語意訊息。但是實際的資料表達方式或儲存方式不可能直接把這些訊息完全的儲存在各別詞項底下。一來是資料量過分龐大，二來是可以有更經濟的處理方式。表示詞彙特性的一個精簡方法是詞彙分類，同一類的詞彙具有相同的語法特性，因此在我們的語法模式底下仍然有詞彙語法的類別。有關中文語法類別可參看 [趙80]。所以實際上每個詞項下的語法訊息，事實上只要標出這個詞的分類訊息及其與同類相異的特性即可。以動詞‘勸’為例，‘勸’屬於 *Vnp, vp* 類，而 *Vnp, vp* 類有如下的特徵：

- (13) 語意訊息
- | | | | |
|-----|------|-------|---|
| 論元： | 主事者： | 意義： | ？ |
| | 對象： | 特徵： | ？ |
| | 題旨： | 意義特徵： | ？ |
- 附加成分：時間，地點，法相，副詞修飾語，語尾助詞，介詞片語
- 語法訊息
- | | |
|------|--|
| 分類： | $V np, vp$ |
| 線性律： | 主事者 $[NP]$ < * < 受事者 $[NP]$ < 命題 $[VP]$
時間 < 介詞片語 < 副詞 < * < 語尾助詞 |

因此屬於‘勸’的個別訊息只有：

- (14) 語意訊息
- | | |
|----|-----------|
| 論 | 特徵：言談類 |
| | 主事者：特徵：人類 |
| 對象 | 特徵：人類 |
| 題旨 | 特徵：動作 |
| 元 | 論元：主事者：□ |

因而每個詞只要儲存分類及其個別差異即可。事實上語法的類別 [張 89] 和語意的類別 [Chen 88] 都形成了一個層次結構 (Hierarchical structure)，這個層次結構是由 is - a 關係所形成，因此具有特性承襲的性質 (Inheritance property)。譬如說動作雙及物動詞，承襲了動作動詞的特性，動作動詞承襲了動詞的特性，動詞又承襲了詞集的特性。在語意方面也是一樣，人承襲了動物的特性，動物承襲了實體的特性，實體承襲了物的特性。

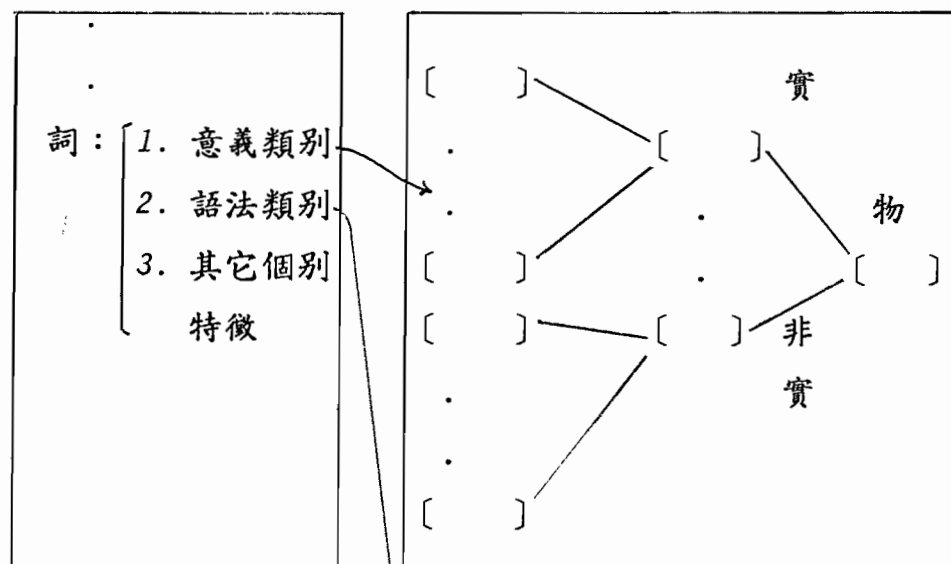
所謂特徵的承襲，事實上就是把高層次的訊息聯併給低層次的子孫。有時爲了表達上的方便，低層次的訊息和高層次訊息有所抵觸時，以低層次的訊息優先。例如，鯨魚是哺乳類但是它沒有一些哺乳類的特徵，例如全身披毛。

詞庫裡每個詞項的資料分佈在三個不同的資料結構中，如 (15) 所示：

(15)

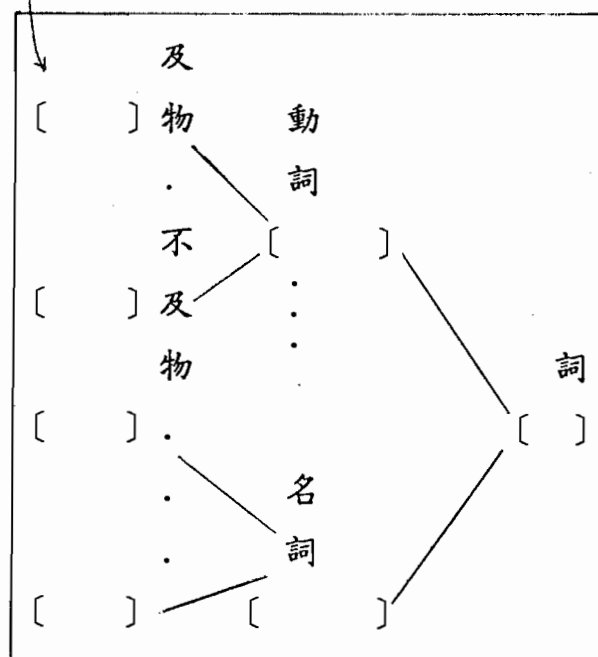
詞檔

語意層次結構



詞： []

語法層次結構



詞檔的資料包含所有的詞彙及其語法分類，語意分類和個別特徵。從語意層次結構檔案承襲了結構層次上的語意特徵，從語法層次結構檔案承襲了語法的分類特徵。將兩種分類特徵和個別特徵聯併以後就得到屬於詞彙的所有訊息以便剖析之用。因此詞彙分析 (*Lexical Analysis*) 的作用除了斷詞之外就是負責從不同檔案組合詞彙的訊息，若有需要並進一步用詞彙規律，得到變化結構的語法訊息。

3.2 剖析與聯併

剖析的過程可以看成是語意訊息的聯併，因為我們假設詞結合為片語時具有意義上的結合性*。語法訊息並不聯併，它的作用只是在導引正確的語意訊息聯併。因此當詞彙組合成片語後，片語的訊息只保留每個角色的語意訊息，語法訊息的部分只保留中心語的語法類別。

聯併的過程就是想辦法填滿必要論元，及找出附加成分。而填入論元時必須依照線性律的規範進行之。

利用訊息為本的格位語法所設計出的剖析器雖然尚未完成，但是從理論及觀察可以預先對他可能的效率和其它語法模式的剖析方式做一番比較。此一語法模式在剖析效率上最大的優點在於語法表達上的精確性與聚焦性 (*Focusing*)，所謂聚焦性指的是有關的語法規律完全在詞彙訊息上，不必考慮無關的語法訊息。例如從線性規律中知道有關某一角色的線性規律只有一兩條，比較之下，可能是從數百條無語境語法律聚焦而得的。另外由於表達的精確性及語意訊息的使用，使得剖析時的分歧路徑 (*Branching*) 減少很多。同時由於角色採用訊息為本的表示法使得詞的多重分類現象減至最少 [陳 89]。另外一個最大的特色是語意分析也同時完成。

* 因此意義上沒有結合性的詞組構造，必須用特殊的方法處理 [Huang 88]，使得聯併時得到有結合性的結果。

例如：1. 吃—NP—豆腐
2. 馬友友的大提琴拉得很好。

四. 結論

這個語法模式的聯併方式和一般聯併語法不同的地方在於聯併的形式較富有彈性，以填空的方式為之，不必考慮每一個特徵路徑 (Feature path) 是否正確對應，減輕許多可能因形式上的錯誤而產生的剖析失敗，不過聯併語法的各項優點依然保留在本語法模式。例如：

1. 宣稱性 (Declarative)

從詞彙上規定什麼是合法句子，和處理程序無關。因此詞彙訊息有所變動時，不必更改剖析程式。

2. 和剖析的運算方式無關 (Algorithm-free)

因此不管採用何種運算方式，甚至採用經驗法則改變執行的順序，都不影響正確結果的取得。

3. 適用於平行處理 (Parallel processing)

由於聯併的過程和執行順序無關 (Order-free)，所以最適合於平行處理。

4. 容許部分剖析 (Partial parse)

聯併是訊息的累積，給予多少訊息就反應多少訊息，因此片語，詞串剖析後得到片語結構，部分詞串得到部分結果，錯誤詞串得到正確部分的結果。

本語法模式除了以上的優點外，還有下列特殊的優點：

5. 語意成分具有普遍性 (Semantic structure is universal)

由於分離語法成分及語意成分的關係，語意成分和語言無關。因此語言的不同是表現在語法訊息上，但是語意訊息是相同的。

只要找到一個良好的中間語言（現在採用中文），機器翻譯就是剖析後做產生的工作。

6. 剖析的結果就是語意表達形式 (Semantic representation)

語法的分析和語意分析結合在一起，不必另有語意分析的過程。

7. 正確的角色定位

語意角色的定位和動詞、名詞本身、常識及標記有關，本語法模式兼顧了所有可能的情形。

8. 剖析系統設計容易，維護容易

由於剖析過程類似於填空 (*Slot and filler*) 方式，因此程式不致於太複雜。

9. 符合人類認知現象

語言能力的表現在於更能掌握詞彙的使用，本語法模式正好反應了這個現象。

10. 具有產生能力 (*Generation*)

詞彙語法訊息，表現了所有可能的合法組合，因此語句的產生只要選擇一組線性規律，符合這組線性規律的詞彙，任何排列即是合法句子。

* 本論文的研究得國科會中文語句剖析的語法模式研究計畫 (NSC77-0408-E001-01) 及中央研究院與工研院電子所合作「中文詞知識庫第三期及中文剖析系統合作研究發展計畫」之部分經費贊助，特此申謝。論文寫作的部分例子取材於中文詞庫小組的研究成果，並感謝詞庫小組其它成員細心的閱讀本文並提供寶貴的意見。

五.參考資料

- 趙元任著，丁邦新譯，1980，中國話的文法，香港中文大學出版。
- 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，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。
- 陳克健，黃居仁，1989，“中文詞的分類與語言剖析”，電腦季刊。
- Bresnan, J. (Ed.) 1982. *The Mental Representation of Grammatical Relations*. Cambridge, Mass.:MIT Press.
- Chen, K. J. , C. R. Huang, and L. P. Chang 1989.
"The Identification of Thematic Roles in Parsing Mandarin Chinese." In *Proceedings of ROCLING II*. Taipei: Computing Center, Academia Sinica
- 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.
- 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. et al. 1987. *Category Structures*. CSLI report 102, Stanford: CSLI.
- Gazdar, G., E. Klein, G. K. Pullum, and I. A. Sag 1985. *Generalized Phrase Structure Grammar*. Cambridge: Blackwell, and Cambridge, Mass.: Harvard University Press.

- Gunji, Takao 1986. *Japanese Phrase Structure Grammar*.
Dordrecht: Reidel.
- Huang, C. R. 1988. "A Unification-based LFG Analysis of
Lexical Discontinuity." *The Fourth International
Workshop on Asian and Oriental Linguistics, Paris*.
- Kay, M. 1985. "Parsing in Functional Unification Grammar." In
D. Dowty, L. Karttunen, and A. Zwicky (Eds.), *Natural
Language Parsing*. Cambridge: Cambridge University Press.
- Pollard, C. and Ivan A. Sag 1987. *Information-based Syntax
and Semantics Vol.1 Fundamental.*, CSLI Lecture Notes
Series No. 13. Stanford: CSLI.
- Sells, P. 1985. *Lectures on Contemporary Syntactic Theories*.
CSLI Lecture Notes Series No. 3. Stanford: CSLI.
- Shieber, S. 1986. *An Introduction to Unification-Based
Approaches to Grammar*. CSLI Lecture Notes Series No. 4.
Stanford: CSLI.
- Uszkoreit, H. 1986. "Categorical Unification Grammars." In
Proceedings of 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.
- Woods, W. A. 1970. "Transition Network Grammars for Natural
Language Analysis." *CACM* 13:10, pp.591-606.