

TR-86-006

中文語句分析系統的研究—語法及剖句

陳克健·林隆基·陳正佳

中研院資訊所圖書室



3 0330 03 000058 7

書	考	參
信	評	不

0058

第一章 引言

本論文針對中文的文法現象及使用計算機分析中文句子的方法做一個初步的介紹。中文語法及其分析法尚待語文學家、心理學家、計算機專家共同努力，以求得更完整更深入的結果。這篇論文只是一個起步，我們整理一套可以用來分析白話中文語句的文法結構，並討論其在電腦處理上的適用性。同時我們也提出了一個語句分析系統模式，這個系統模式考慮了中文語法的特徵及計算速度，彈性設計等原則而完成的。

做這個研究的目的，是為電腦了解自然語言（中文）做一個鋪路的工作。電腦了解自然語言必須做句子語意的分析，而語意的決定和語法有密切關係，語法結構決定一句話內什麼是主語，什麼是賓語。當然如果分析自然語言的結構完全依靠語法也是不行的，有的時候語意的分析可以解決一些語法上的含糊 (Ambiguity) 現象，例如：

① 打倒魔鬼的英勇武士。

② 打倒魔鬼的英勇武士。

從語意上看②的結構不合理，但從語法上看①、②都完全合法。因此，語句的分析應該考慮採用語法和語意並行的分析方式似乎是比較合理的方式。本篇論文尚未將語意的分析包含在內。希望從純粹的語法分析當中，讓我們能先了解一些語法現象及語法處理上力有未逮之處。

1.1 語法與語句分析之定義

一個中文句子可以視為一串中文字及標點符號所構成，例如：

(1) 張三在公園裏看到李四。

(2) 大明，昨天被他媽媽打了一頓，今天離家出走了。

中文句子有比(1)句，更簡單，也有比(2)句更複雜的，但是這些句子都有規則性，並非任意組合中文字就可以成一個句子。有些不合文法的句子，我們可以了解它的意義，但它不是一個合法的句子。例

(3) * 很痛，手心打

大概是，手心被打得很痛的兒語，這個例子又說明一件事，語意的分析和語法不是絕對有關的，了解一個句子並不必要知道完整的語法結構。應如何彈性的處理語法及語意現象，是值得將來深入研究的題目。在這裏我們假設一個合法的句子，有固定的文法結構。

一個句子的文法結構，因不同的文法論，不同的分析方法，而會有許多不同的看法，目前，我們只舉出一種最通常被使用的

直接成份 (immediate constituents) 表示法：

這種結構表示法是把一個詞組的結構，看成是由這組的成分依序排列而成。例如：一個句子可能是由一個名詞片語後接一個動詞片語組合而成，而一個動詞片語又可能是由一個介詞片語、一個動詞、一個名詞片語依序排列而成。例如：以 張三在公園看到李四而言，這個句子的可能結構如下圖：

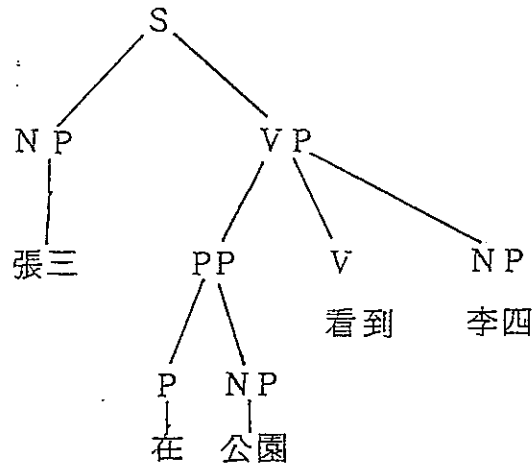


圖 1.1

所謂語句分析，就是把一連串的中文字，經過分析得到樹狀文法結構的過程。在這裏我們把語句分析的過程分成兩個步驟，如圖 1.2 所示。

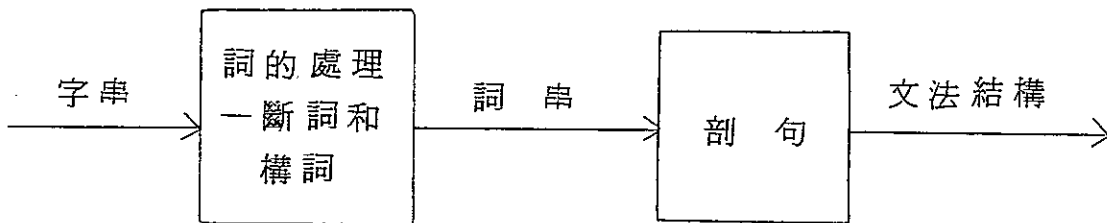


圖 1.2

第一個步驟一詞的處理，我們在〔1〕中已經討論過了，這裏只補充說明一點，從實用的觀點來看，把析句分成兩個步驟是很自然也很合理的，但是在第一步構詞的過程中字的組合成詞和剖句中詞組的組合，似乎有相同的特性，及部分相同的處理方式，因此兩段式的處理及劃分方式並非絕對如此。

本文主要的討論重點放在剖句及其使用的模式及文法上。

第二章 中文語法綱要

以下所討論的資料，取材於〔2,3,4,5,6〕，我們將適用於電腦處理的部分綜合整理，抽出主要部分，如詞類及片語，句子的結構。讀者如要進一步了解中文的文法現象，除了〔2,3,4,5,6〕之外，〔7,8〕有較詳細的探討。

2.1 中文語法特徵

中文跟其他語言比較起來，具有下列重要明顯特徵：

1. 中文沒有表示格 (case) 的詞素，很多語言都具有詞素可以表示名詞在句中的文法功能，但是在中文中，一個名詞在句中的文法功能主要是藉詞序、語義，或者介系詞來表示，而不是用不同型態的名詞來表示。
2. 中文沒有表示名詞為單數或複數的記號，雖然中文中有些名詞可藉附加詞尾“們”，以表示它是複數，但是這只有代名詞和少數具有人的屬性的名詞，才可以在大部分情形下，每個中文名詞都可表示單數或複數。
3. 中文沒有動詞、名詞間必須一致 (agreement) 的限制，很多語言在遇到不同人稱，或不同數 (單 / 數) 的名詞擔任主語或賓語的時候，就必須使用不同型態的動詞，這種動詞名詞間必須“一致”的限制，在中文句子當中是不必要的。不過中文卻有分類詞，名詞間必須一致的限制，例如：我們可以說“一本書”，但是卻不可說“一條書”或“一個書”。

4. 中文動詞只有時貌 (aspect)，而沒有時式 (tense)：

時式指的是事件發生時間與說話時間的相對時間關係，而時貌指的則是表示某一事件相對於另一事件是在持續中或已完成。例如(1)中

(i) a. She is walking

b. She walked

a 句是現在式和進行貌，而 b 句只有表過去的時式，而無時貌。

國語沒有表時式的記號，只有表時貌的記號“了”表“完成”

，“過”表“經驗”，“著”表“持續”。

5. 中文沒有明顯而規則的詞類變化，在英文中常常可找到一群形態不一樣，但語意相近的詞語分別在句中擔任不同的文法角色。

例如：beauty, beautiful, beautifully, beautify, 這四個字詞意近似，他們分別在句中擔任名詞、形容詞、副詞與動詞。但是在中文裡不同功能的文法角色，往往都只由同一個詞語擔任。

6. 中文是一個主題明顯 (topic-prominent) 的語言，討論中文句子結構時，單是主語、賓語等，這些概念並無法完全表示出某些名詞片語在句中的文法功能，中文還需要“主題”的概念，一般而言，任何語言的句子，我們大致都可將它分開成二個語法部分，一個是主語，一個是謂語，在某些語言裡（例如：英文）這種主謂關係可以被解釋成爲動作者—動作的關係，亦即主語必須跟表示動作和狀態的動詞有直接的語意關係。例如：

(1)(a)我已經回來了

(b)張三長得很健壯

上述中的“我”“張三”都可算是主語

(2)(a)張三肚子很大

(b)張三他已經請過了

但是在(2)(a)中，“很大”的是“肚子”，而不是“張三”，因此主語是肚子而不是張三，張三在句中的角色變得不明。

在(2)(b)中，“請過”這個動作的動作者是“他”，因此，“張三”不是主語，“他”才是主語。

由於像(2)式的句子，在中文裡常常出現，它是動作者—動作式的解釋法無法說明清楚的。因此，如果也採用主題—評論式的解釋（亦即把主語視作主題，把謂語視作是對主題的評論），將比較能夠分析清楚某些中文句子的結構。

7.中文詞序特徵主要有：

(A)關係子句，所有格片語都出現在主體名詞之前，如(1)：

(1) a. 張三看到的東西

b. 我的背包

(B)時貌記號跟在動詞之後

(2) a. 我去過台北

(C)助動詞出現在動詞之前

(3) 它可以來了

(D) 除了時間和處所片語外，介系詞片語多出現在動詞之前

(4) 老李在家裡看電視

(E) 主受詞都可能出現在動詞之前的任何位置

(5) a. 張三電影看過了

b. 電影張三看過了

以下我們就中文詞組的類型一一介紹。

2.2 名詞片語 (Noun Phrase)

一個名詞片語的構成，主要是名詞和它的一些修飾語所組成。在中文裏這些修飾語都必須在名詞前出現，如分類 (classifier) 度量詞片語 (measure) ，連合片語 (association) ，形容詞，關係子句 (Relation Clauses) 等。也有一些名詞通常是不加修飾語的，如專有名詞 (proper nouns) ，代名詞 (pronouns) ，地方名詞 (place words) ，時間 (time words) 等。

名詞片語在一個句子裏通常扮演的角色為主語、主題、賓語，或介詞的受詞。下面就是一些名詞片語的例子：

(a) 名詞：人，火，觀念，地方。

(b) 專有名詞：李白，愛迪生。

(c) 代名詞：他，誰，什麼。

(d) 地方名詞：台北，歐州。

(e) 時間：昨天，今年。

(f) 位置詞：前面，上，東方。

(g) 有修飾語的名詞片語：這張三十斤重的桌子，張三昨天買的

那幾本很好看的武俠小說。

我們現在討論名詞、修飾語的結構

2.2.1 數量 / 分類詞片語 (QP)

$$\text{QP} := \left[\text{DET} \right] \text{NO.} \left\{ \begin{array}{l} \text{CL} \\ \text{MS} \end{array} \right\}$$
$$\left| \begin{array}{l} \text{DET} \\ \text{MS} \end{array} \right\} \left\{ \begin{array}{l} \text{CL} \\ \text{MS} \end{array} \right\}$$

在這裏我們採用了一個表示文法結構的表示法，:= 表示左邊的東西是由右邊的組合而成；[] 表示可有可無；{ } 表示必選，| 表示或 (OR)。

表式中 CL 表分類詞，如 "張"，"個"，"條"，"隻"。

MS 表度量詞，如 "斤"，"瓶"，"公斤"。

DET 表定詞，如 "那"，"這"，"每"。

NO 表數詞，如一、三、二十幾、數百。

QP 放在名詞前修飾名詞，形成名詞片語如："三本書"，"那幾個人"，"十兩肉"。度量 / 分類詞也可獨立做爲代名詞，例如：那些蘋果請給我幾個。

2.2.2 連合片語 (association phrase)

當兩個名詞片語用 "的" 連接時，前面的名詞片語和 "的" 構成了一個連合片語，其公式如下：

(1) NP 的 NP

連合片語 主體名詞

連合片語與主體名詞之間，連合的關係完全決定於各別名詞片語

。其中有一類的連合片語即屬所有格片語，如：

- (2) a. 張三的褲子
- b. 他們的房子
- c. 牛的胃

其他非屬所有的連合片語有：

- (3) a. 學校的宿舍
- b. 校園的風景
- c. 夜晚的天空

連合片語的另一類亞型是：

人稱代名詞＋親屬稱謂名詞，例如：

- (4) a. 我看過你哥哥
- b. 她姊姊嫁人了

2.2.3 修飾片語：

修飾片語主要置於名詞之前，以修飾主體名詞，它包括了關係子句與屬性形容詞，關係子句的字序如下：

(1)關係子句：＝子句＋“的”

 | 動詞片語＋“的”

關係子句的語意動態在於更進一步說明主體的功能，屬性形容詞則是指不加“的”，直接用來修飾主體名詞的形容詞，如：

- ① a. 假話
- b. 天然顏色
- c. 白花

名詞片語

名詞片語的成份次序如下

- ② a. 連合片語 + 分類 / 量度詞片語 + 關係子句 + 形容詞 + 名詞
- b. 連合片語 + 關係子句 + 分類 / 量度詞片語 + 形容詞 + 名詞

例如：② a. 做生意的 那個商人

關係子句 QP

b. 我那個 住在美國的 好朋友

AP QP RLC ADJ

c. 那個 飯店的 菜

連合片語

這個公式並不完全，但爲了增進系統的效率，我們目前只接受到此。

例如：“昨天來的那個我不喜歡看的人”，在此句中，可以同時有二個關係子句。

2.2.4 形容詞

形容詞可以修飾一個名詞，有時候也當成謂語。我們在下一節動詞片語中會討論到。

形容詞加名詞形成名詞片語，例舉如下：好孩子，笨學生，紅花，但是大部分的形容詞修飾名詞時中間要加一個“的”。

(a)舒服的椅子

(b) * 舒服椅子

(c) 紅紅的花

(d) * 紅紅花

2.2.5 名詞修飾語的順序

一個名詞片語，可以是很複雜的形式如：張三昨天買的那幾本很好看的金庸的武俠小說。

這種例子在實際生活上很少見，以下幾類最常見：

(a) S 的 + QP + ADJ + N

(b) NP 的 + QP + ADJ + N

(c) QP + S 的 + ADJ + N

(d) QP + NP 的 + ADJ + N

事實上，計算機處理這類名詞片語的時候，修飾語的順序並不很重要，只要知道那些字是名詞的修飾語即可。不過一般的原則是沒有加“的”的形容詞最接近名詞。

2.2.6 位置詞 (localizers)

位置詞可以看成是名詞的一種，例如，火車站的前面，那個人的後面，前面有許多人。

但很多時候它出現的形式是：

NP + 位置詞

例如：(a) 桌上

(b) 火車站前面

(c) 我住的那棟房子裡

許多位置詞也可視為地方名詞，例如：前面有許多人，房子的左邊。

2.3 動詞片語 (Verb Phrases)

在一個句子中扮演謂語 (predicate) 這個角色的都可以看成是動詞。因此有些形容詞在句中扮演謂語的角色，也被視為動詞。如：

(a) 他很高

(b) 我快樂

形成動詞片語的成分，除了動詞之外還有助動詞 (Auxiliaries)，副詞 (adverbials)，受詞 (objects) 及一些相關的元素如時貌、介系詞片語等。我們將在下述小節中一一加以解說。

2.3.1 時貌

中文中有四個時貌用字 "了"，"過"，"著" 和 "在"，它們和動詞連用表示動作和時間的關係。其中只有 "在" 是放在動詞的前面，其它三個都在動詞之後。

例如： 他在洗澡。

我看過那部電影。

2.3.2 置於動詞前的元素 (Preverbal Elements)

有三種不同的詞類可以出現在動詞及主語之間

A. 助動詞：應該，會，可以，敢。

B. 介系詞片語 (PP).

介系詞片語是由介系詞 + NP 組合而成。

如 (a) 他向我借錢。

(b) 他比你胖。

(c) 他從樓上跳下來。

詳細的介系詞分類見 [8,9]。

C. 副詞成分

副詞成分包含：(a)副詞：沒，不，很，高興地，大概，都，在。

(b)時間副詞：昨天，這個學期。

2.3.3 置於動詞後的副詞成分

大部分的副詞成分都出現在謂語的前面，只有少數五種是出現在動詞的後面的，而且絕對不可以出現在動詞前。它們的形式有下面兩種。

(a) V + QP

(b) V + 得 + Clause

例如：

(1)表次數的補語：

(a)他來過兩次

(b)他被踢了一脚

(2)表持續時間的補語：

(a)他睡了三個鐘頭

(b)我等了十年

(3)表數量的補語：

(a)他比你胖十公斤

(b)他跑了十公里

(4)表結果的補語：

(a)他跑得很累

(b)他吵得我受不了

(5) 表量測的補語：

(a) 他跑得很快

(b) 他吵得很兇

2.3.4 動詞的分類及動詞片語的形成

語言教學上把動詞分爲不及物、單賓、雙賓動詞。從計算機處理自然語言的觀點來看這樣粗略的劃分是不夠的，一般而言，動詞幾乎是決定句子結構的主要因素，會生成同樣結構的動詞應劃分爲同一類，計算機程式就可以依據不同的動詞分類去分析句子的結構。在這裏我們應實際的需要，把動詞分爲下列幾種類型。

(1) 形容詞 (ADJ) 例：重，高興

(2) 不及物動詞 (V-) 例：死，來

(3) 及物動詞 (V-n) 例：打，喜歡

(4) 雙及物動詞 (V-nn) 例：送，給

(5) 受詞爲一個句子的動詞 (V-s) 例：說，相信

(6) 受詞爲動詞片語的動詞 (V-v) 例：設法，願意

(7) 間接受詞爲名詞片語直接受詞爲句子的動詞 (V-ns) 例：告訴，報告

(8) 直接受詞爲名詞片語而以動詞片語爲受詞補語的動詞 (V-nv)

例：叫，請

(9) "是"

(10) "有"

以上動詞分類的詳細說明及其例句見 [7]。

動詞片語 (VP) 的結構如下：

$$VP := [AUX \mid PP \mid ADV \mid NP]^* V'$$
$$V' := V' [QP \mid 得 S]$$
$$V' := V [NP] [NP \mid PP \mid S]$$

然而剖句過程中，動詞片語的分析將依據動詞的分類爲之，例如：

他送我一本書。

剖句程式見到動詞“送”，即查驗知“送”屬V-nn 類之動詞，因此利用 VP:= 送 NP NP 或 VP:= 送 NP給NP 爲剖句的文法。

2.4 句子

簡單的說一個句子是由一個名詞片語加動詞片語組合而成的。有時候，最後可以再加一個句尾助詞 (sentence-final particle)，如“了”，“吧”，“嗎”，“呢”，“呀”。這些句尾語助詞，除了“了”之外都必須加在一個句子的結尾。

中文句子的第一個名詞片語不見得是句子的主語，例如：

(1) 這種樹 葉子很大

(2) 那場火 幸虧 消防隊 來得早

例子中“葉子”，“消防隊”才是主語，“這種樹”，“那場火”，只是句子的主題 (Topic)。主題的角色很廣，可以是句子的主語、賓語、補語等。

一個句子甚至於可以沒有主語，只有主題的情形，如：

(3) 房子蓋好了

(4) 功課做完了沒有？

從語法結構上，我們沒有辦法區別主題和主語，我們把這個問題視爲語意分析的範疇。

2.4.1 疑問句

中文有四種不同的疑問方式

(1) 疑問詞問句

你喜歡誰？

(2) 選擇式的問句

你喜歡他還是喜歡我？

(3) 句尾語助詞及附加問句

你喜歡張三嗎？

你喜歡張三，對不對？

(4) 是非問句

你會不會唱？

你喜不喜歡張三？

分析疑問句和直述句並無任何差異。在第(1)類中疑問詞可以當作代名詞處理。第(2)類中“還是”是連接詞。第(3)類句尾語助詞及附加語永遠出現在句尾，可以從句子中除去。第(4)類的是非形態動詞如會不會、喜不喜歡可以把它當做單一動詞來處理，“會不會”看成“會”，“喜不喜歡”看成“喜歡”即可，句子的結構並不因動詞成 X- 不 -X 而有所改變。

2.4.2 主語反置現象

除了被動語態，主語的位置總是出現在動詞的前面，鮮有出現在動詞後面的情形，下面就是少數主語反置的情形，例如：

前面坐著一個人

“坐著”本來屬於不及物動詞，在這個例子裏，語意上的主語好像

是“坐著”的受詞，位置副詞成了主詞。按不及物動詞的語法分析，這個句子根本不合文法。因此令我們更覺得語法的分析必須和語意分析並行。

2.4.3 句子的語法結構

從以上的討論，我們可以把句子簡單的歸納成下面的幾種結構：

$S'' : = [\text{Topic}] S'$

$S' : = S [\text{PRT} [\text{TAG}]$

PRT : 助詞

$S : = [\text{NP}] \text{VP}$

TAG : 附加

在 2.3.4 節中動詞分類中 V-s 及 V-ns 型動詞以句子做爲直接受詞時，這裏的句子指的是最廣義的 S'' 型，例如：

我以爲那本書張三看過了。

Topic

第三章 剖句的方法

一般而言剖句的方法分爲兩種，一種是由上往下 (Top-down) 另一種是由下往上 (bottom-up) 的方法。各有利弊。由上往下的方法最大的缺點，就是可能會產生很多回索 (Back tracking) 現象，猶其中文文法，許多不同的組合，可以形成相同的語法成分。如果分析時選錯組合的規則，會浪費很多時間在回索上。如果採用由下往上的分析法，又會在分析過程中組合一些不必要的語法成分。因此我們採用了一種取雙方之長的圖表法 (Chart parser) [7,10]。

3.1 圖表 (Chart)

所謂一個圖表，像是一個圖形，這個圖表中每一條徑 (edge) 代表了一個語法成分 (constituents)。在分析過程中不斷有新的語法成分被合成，在圖表中就增加一個新的徑標明其成分名稱。這些徑的用處除了可以紀錄已完成的部分工作之外，並可以幫助決定如何去產生下一個新的語法成分。圖 3.1 是一個句子的起始圖表，這個起始圖表是由斷詞部處理後產生的結果。

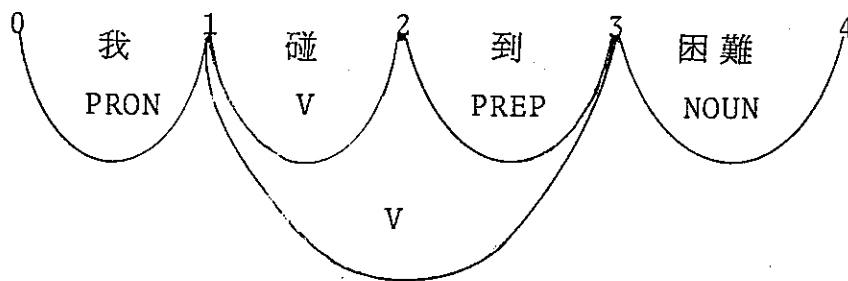


圖 3.1 一個句子的起始圖表

從圖表中的徑，我們知道那些分析已經完成不再重覆之外，還利用這些已完成的徑去決定如何尋找下一個語法成分。回憶在 2.3 節中我們把動詞依組成句子方式不同而分類，其目的是希望做到唯一分析 (deterministic parsing) ，在圖表式的分析過程中我們碰到了一個語法成分的主要部分 (Head) 時，例如圖 3-1 例句中的動詞 " 碰到 " ，就開始起動一個新的徑的合成 (active edge) ，我們知道 " 碰到 " 是屬於 V-n 類，因此馬上開始尋找 V-n 類動詞，形成動詞片語 VP 尚缺的 NP 如果用圖形表示就如同圖 3.2 。

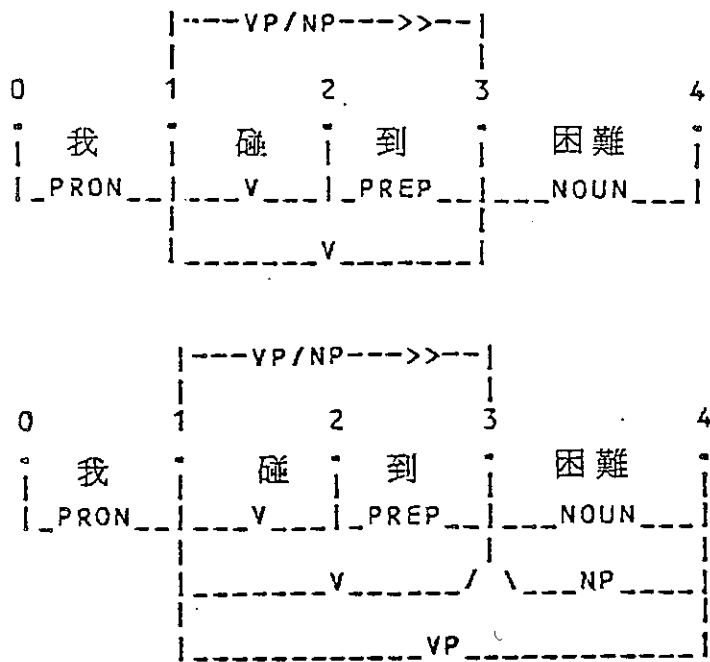


圖 3.2 起動徑 VP/NP----->> 的產生及結合NP後之結果

這個NP又知道在動詞的右邊，因此在 " 碰到 " 的右邊找到一個 Noun的徑，可以變成 NP 於是 " 碰到困難 " 就形成了一個 VP, 在這裏有一點要提的就是這個被起動的徑 VP/NP--> 並不因此而消失，因為可能這一次找出來的VP成分並不正確，也許有其它可能的組合，這種趨向唯一性質的分析法，能夠成功主要還是依靠，詞類劃分時

做得較為細緻，因此名詞及動詞分類在自然語言分析上是非常重要的工作。

實際上一個圖表上所包含的資料，要比圖形上顯示的多。目前徑上包含了5種不同的資料。(1)語法成分。有兩種不同的徑，一種是已找到的語法成分，另一種就是起動的徑(Active edge)。(2)語法成分的其它分類資料(3)組成一個徑的成分(4)語法及語意特性。例如動詞"吃"的主詞必須是動物等。(5)圖形的指標，如每一個徑的左邊點及右邊點。

3.2 圖表式的分析法

下面是圖表分析法的程序及規則：

(1)起始圖表由斷詞部產生。斷詞部把輸入的字串斷為詞串並將，每一個詞的語法分類及特性資料填入。

完成(1)後，重覆執行(2)，(3)，(4)直到分析成功或沒有新的語法成分可以產生時宣告失敗。

(2)對每一個語法成分形成的徑，找其相鄰的徑，若有相鄰的起動徑，則執行步驟(5)。

(3)對每一個起動徑，找其相鄰的徑，若存在，則執行步驟(5)。

(4)對每一個語法成分形成的徑，檢查語法規則，若這個徑是某一規則的主要成分(Head)，則產生一個起動徑。

(5)當有一個起動徑和相鄰的語法成分徑被找到時。

(5.1)若能結合成一個新的語法成分，則製造一個包含這個語法成分的新徑。

(5.2)若不能組成新的語法成分，則不做任何異動。

(5.3) 若兩者不能組成新的語法成分，且起動徑本身為一個完全的語法成分，則更改此一起動徑為，語法成分徑。

從以上的步驟(4)，我們知道語法成分應分為兩類一類是主要成分，一類是組成分。像V'是VP的主要成分，N（包含PRON,NOUN等）是NP的主要成分，V是V',VP是S的主要成分。每一個組合律中都有一個主要成分。

第四章 結論

由於中文文法的特性與其它語言有相異之處，我們在第 2.1 節中已有詳細的說明。因此採用的分析方法及文法的表達模式並沒有採用 ATN [10,11] 形式。ATN 採用由上到下的分析法對結構自由度比較大的中文並不適合，同時利用共用變數傳遞特徵資料來完成時式、時態、語態，及人稱，單複數一致的方法，對中文文法分析並無太大幫助。我們所採用的圖表式的分析法既簡單，又有效率。實際用 LISP 語言在 VAX11/780 完成實驗結果顯示，一般句子除去斷詞部分可在約 0.1 秒內完成剖句的工作，下例是一個比較複雜的句子，花了 0.118 秒。

%INPUT ==> 這種錯誤將來也許會被大家認為是對的。

%Parsing time: 118 ms

(SENTENCE

(S

(NP <NOUN>

(QP <Mc>

(DET ==>> 這)

(CL <Mc> ==>> 種))

(N <NOUN> ==>> 錯誤))

(VP

(ADV ==>> 將來)

(ADV ==>> 也許)

(AUX ==>> 會)

(PP <bei>

(PREP <bei> ==>> 被)

(NP <NOUN>

(N <PRON> ==>> 大家)))

```

(V-bar
  (V <V-s> ==>> 認為)
  (S
    (EMPTY (No.1))
    (VP
      (V-bar
        (V <SHI> ==>> 是)
        (XPDE <S>
          (S
            (EMPTY (No.8))
            (VP
              (V-bar
                (V <ADJ> ==>> 對)))
              (DEn ==>> 的)))))))))

```

%EMPTY.8 == EMPTY.1
 %EMPTY.1 == NP:(這種錯誤)

這個實驗系統並沒有解決所有的中文剖句問題。在此我們提出一些猶待解決的問題，大部分的情形都是由於結構上的含混而引起的。

1. 主題和主語的含混

(1.a) 功課做好了。

(1.b) 我做好了。

這兩個句子都是NP+V的形式，但實際分析句子結構時卻大不相同。句(1.a)中功課是主題而非主語，分別主題及主語時要用語意

2 數量詞片語的修飾範疇 (Attachment of QP)

(2.a) 一個小孩子的書 [[[QP NJ 的] N]

(2.b) 一本小孩子的書 [QP [N 的] N]

(2.c) 那罐小狗吃的食物 [QP [N V 的] N]

(2.d) 那隻小狗吃的食物 [[[QP N] 的] N]

解決這個問題必須把數量詞及所能修飾的名詞種類放入字典內。

3 名詞片語的含混

3.1 所有格結合性的問題

(3.a) 他的爸爸的朋友 [[N 的 N] 的 N]

(3.b) 他的金庸的小說 [[N 的] [N 的] N]

(3.c) 打倒魔鬼的武士 [V [[N 的] N]]

(3.d) 打倒魔鬼的武士 [VP 的 N]

所有格結合性的問題，也要依靠語意才能解決。

3.2 '的' 的省略

'的' 在代名詞後可以被省略，N-N 複合名詞被省略。

(3.e) 我看過你(的)那本書

(3.f) 大理石桌子，水果小販

4 多重詞類的含混

'制服' 可以當名詞解，也可以當動詞，兩者語意完全不同。

'代表' 為名詞也為動詞，但語意是關聯的。

5 文法上的含混

相同的組成分子，可以產生不同的句型結構

(5.a) 老師叫我們上課的時候(我們)不要說話。

(5.b) 老師教我們跳舞的時候(老師)摔了一交。

這些含混的問題大部分可以用語意解決，因此如何建立一個語意與語法整合的分析方法，是一個主要研究的課題。

參考資料

- 1 陳克健、陳正佳、林隆基，「中文語句分析的研究—斷詞及構詞」，中研院資訊所技術報告，1986.
- 2 趙元任，中國話的文法，台北，學生書局，1970.
- 3 黃正德，「中文生成語法概要」，中研院資訊所演講稿，1984.
- 4 Li, Y. C. and Cheng, Robert L. et. al., Mandarin Chinese: A Practical Reference Grammar for Students and Teachers, Vol. 1, Crane book co., Taipei 1984.
- 5 Li, Charles N. and Tompson, Sandra A., Mandarin Chinese: A Functional Reference Grammar, Berkeley, University of California Press, 1981.
- 6 湯廷池，國語語法論集，台北，學生書局，1981.
- 7 林隆基，「中文語法分析系統」，國立台灣大學電機工程研究所，碩士論文，1985.
- 8 陳正佳，「一套中文語法分析系統的研究與設計」，國立台灣大學資訊工程研究所，碩士論文，1985.
- 9 Li, Charles and Thompson, Sandra, "Coverbs in Mandarin Chinese: Verbs or Prepositions?" JCL 2.3., 1974.
10. Winograd, T., Language as a Cognitive Process, Vol. 1: Syntax, Addison-Wesley, 1983.
- 11 Woods, W. A., "An Experimental Parsing System for Transition Network Grammars", In: Rustin ed. Natural Language Processing, Algorithm Press, N.Y., 1973.