

# 漢語的時間詞組和語言剖析

張莉萍

中央研究院計算中心，輔仁大學語言所

## 論文摘要

時間詞組在漢語中所扮演的角色及出現的位置，很容易讓電腦在剖析 (parsing) 語言時，誤以為是動詞的必要論元 (argument)，而誤導剖析的方向或使得剖析的過程複雜化。因此，本文整理出漢語時間成分的各種表達形式並提出判定時間成分的模式，使電腦能辨認時間詞組並解決含混情形。此外，本文試圖為漢語的時間概念分類，希望從它們的形式特徵以及出現在句中的位置幫助電腦進一步的了解句意。

## 1. 前言

剖析漢語時，動詞是整句的中心，適當的動詞分類可以決定一個句子的語法結構及相關的語意訊息 [Chen et al. 88]。因此，規定一個動詞帶幾個論元 (argument)，對於電腦剖析 (parsing) 自然語言而言，是非常重要的訊息。但是，有時句子的某些成分 (constituents) 並非動詞支配的語意論元，如：副詞、連接詞、地方詞、時間詞。尤其我們在分析天下雜誌上的文章時，發現幾乎每三個句子中，就出現一次表示時間的成分，出現的頻率相當高，而且時間成分往往是名詞性的詞(組)；不像副詞、連接詞這些成分，可以由詞類得知它們不是動詞的必要論元；也不像地方詞，往往由介詞引介出現；而時間詞組常常因為它的詞性及出現的位置，容易被誤以為是動詞的必要論元，如果電腦能辨認出它們是“額外”的論元，那麼不僅可以簡化剖句的過程，而且可以節省剖句的時間。例如“他寫了三個月，才寫了一篇短文”，這裡的“三個月”不是動詞的必要論元，而是表示“寫一篇短文”這個動作所持續的時間。如果電腦剖析時，就知道“三個月”是時間詞，而且放在動詞後面 [註1]，並不是動詞的必要論元，就不會把它和整個句子做匹配 (matching) 的工作。又例如“昨天他没來”這個句子中，動詞前有兩個名詞組，如果電腦知道“昨天”是時間詞，就可以只管“他”是不是符合動詞的格框 (case frame) 及語意限制 (semantic restriction)，而不用管“昨天”了。可見任何表示事件的語句，都可以加上時間概念來描述補充一個事件，而且在句中出現的頻率又高。本文就從時間成分著手，以天下雜誌中的語句為範例，

針對中文的剖析步驟，整理出漢語中時間的表現方式及其成分，放入辭典(*lexicon*)中，讓電腦在斷詞時能立即判斷出句子中的時間成分，以簡化下一步的句子分析過程，進而幫助電腦理解句意。

## 2. 時間成分的種類及判定方式：

首先，我們在整理漢語語法時，將時間成分依詞性分為三大類：一是時間名詞(組)，二是表時間的介詞片語，三是時間副詞。下面就這三類的內容與語法特徵，分類詳加討論電腦如何判定它們是表示時間的成分。

### 2.1 時間名詞(組)

依構詞法再分為三小類：

#### a. 簡單式時間詞：

這一類由單純的名詞(沒有任何標記)形成的時間詞，在辭典中是列得完的一類，包括：

朝代名，如：夏、商、周、...

年號名，如：乾隆、雍正、...

干支名，如：甲子、丙寅、...

節氣(或節日)名，如：冬至、雨水、端午、中秋、...

副詞性時間詞，如：現在、過去、未來、...

一般時間詞，如：早上、中午、晚上、時代、朝代、日子、時候、時期、月、星期、...

#### b. 定量式時間詞(組)

定詞加時間量詞構成時間詞(組)，例如：“每天”、“上禮拜”、“前三天”[註2]。我們不可能在辭典中窮舉這一類的詞，不過，我們可以窮舉時間量詞以及能和時間量詞構成時間詞(組)的定詞。如下所示：

定詞：數詞定詞、這、那、哪、前、後、上、下、初、周、今、明、昨、去[註3]、第、半、多[註4]、每、某、本、此、該、當、全、整、幾、多少、許多、很多、好多、好些、好幾。

時間量詞：世、紀、世紀、年、年度、載、歲、月、旬、星期、禮拜、週、日、天、夜、時、更、點(鐘)、刻(鐘)、分(鐘)、秒(鐘)、小時、期、學期、季、朝、回、次、陣(子)、會(兒)、生、輩子等[註5]。

另外，我們建議把週(或禮拜或星期)一、二、三、四、五、六、日(或天)及初一... 初十，這些數量有限，構詞法又很特別的時間名詞當做漢語的時間專有名詞放入辭典中。

### c. 和方位詞構成時間詞(組)

這一類可以是名詞組、動詞組或句子加上方位詞形成表示時間的成分，例如：“春天裡”、“消息傳出後”、“他進門之前”。由於方位詞所標誌的可能是表示地方的概念，也可能是表示時間的概念，如：“桌子上”的“上”一定標示地方，“十八世紀時”的“時”一定是標示時間；因此，我們需要將含有時間特徵的方位詞標記出來，以便和具有空間特徵的方位詞識別開來。時間方位詞是可以窮舉的一類，如右：前、以前、之前、後、以後、之後、內、以內、之內、中、之中、間、之間、中間、當中、來、以來、頭、裡、底、末、終、初、起、時、開始、左右。又因為時間方位詞中，有部分也兼空間方位詞 [註6]；就我們歸納的結果發現，這些方位詞前的成分如果是時間詞組或是帶事件特徵(event feature)的詞組(以下簡稱Event)，才能說和方位詞構成的是一個表示時間的成分。所謂帶事件特徵的詞組 [註7]是指名詞組、動詞組或句子蘊含事件的語意概念，如：“車禍”。而和Event共存的時間方位詞有：時、中、中間、當中、間、前、後、以前、以後、之前、之後、以來。例子有：“他當財政部長時”、“休息中”、“談話中間”、“演奏當中”、“慌亂間”、“股市尚未崩潰前”、“消息傳出後”、“他就任以來”。

知道了以上三種構成時間詞(組)的方式，要讓電腦在句子中辨識出時間詞(組)就不是件難事了。我們的作法是在時間名詞、時間量詞的詞項記載(lexical entry)之下給予詞類以及時間特徵(time feature)。詞類是幫助電腦剖析句子時，知道什麼樣的詞類可以結合成一個成分(constituent)，時間特徵是讓電腦能判別它們是表示時間的成分，如(1)“上”和“星期”構成一個時間成分，(2)“下”、“個”和“月”構成時間成分。

(1)	他	上	星期	去	美國。
	N	定	量詞	V	N
			[+time]		

(2)	他	下	個	月	考試。
	N	定	量	N	V
			[+time]		

這些詞組都可以由時間量詞或時間名詞判斷出是表示時間的成分。

如果是 *Event* 和時間方位詞構成一個時間成分，我們的作法是規定它們有特徵共存限制(*feature cooccurrence restriction*) [Gazdar et al. 85]，隱含時間特徵，如下所示：

$$[+Event] \ \& \ [+TLOC] \ \supset \ [+Time]。$$

例如(3)"進去時"是一個帶事件特徵詞組和方位詞構成的時間成分，"戴"才是這個句子的主要動詞。再如(4)，第一個子句是表示時間的附屬子句，第二個子句才是句子的主要子句。

(3) 他	進去	時	戴	了	一頂紅帽子。
<i>N</i>	<i>V</i>	方位詞	<i>V</i>	<i>Asp</i>	<i>NP</i>
	[+Event]	[+TLOC]			

(4) 消息傳出後，大家都很擔心。

## 2.2 表時間的介詞片語

關於表時間的介詞片語這一類，由於介詞所引介的對象不只一種，可能引介時間、地點、工具、情況等語意角色，爲了讓電腦能辨別介詞所引介的是一個表示時間的成分，我們整理分析了所有表達時間觀念的介詞片語，得到了"介詞+時間詞(組)+後續成分"這個句式 (*pattern*)，並窮舉各個介詞和後續成分的特徵限制。茲分爲下列七類，每類下面各舉兩個例子：

- a.  $\left\{ \begin{array}{l} \text{在} \\ \text{於} \end{array} \right\} \left\{ \begin{array}{l} \text{時間名詞(組)} \\ \text{Event} \end{array} \right\} \left\{ \begin{array}{l} \text{前、之前、以前、之後、以後、後、} \\ \text{時、中、間、中間、內、裡、初、} \\ \text{末、底、終、左右、(的)時候。} \end{array} \right\}$

(5) 在這問題未解決之前...

(6) 部分結果終於在十月下旬出來了...

這一類如果省略後續成分，介詞後只能接時間名詞(組)。

- b.  $\left\{ \begin{array}{l} \text{到、直到、} \\ \text{等到} \end{array} \right\} \left\{ \begin{array}{l} \text{時間名詞(組)} \\ \text{Event} \end{array} \right\} \left\{ \begin{array}{l} \text{時、(的)時候、後、} \\ \text{之後、以後、(爲)止} \end{array} \right\}$



外，附屬成分或功能詞所引介的語意角色，則要靠進一步的語法整理來幫助電腦連結或理解語意。然而漢語中的功能詞(即介詞)，通常不只引介一個語意角色，因此，我們整理出來的表示時間的介詞片語句式，可以幫助電腦判斷在什麼樣的情況下，介詞所引介的是一個時間成分。如(20)

(20) 大家趁天黑趕快行動。

"天黑"是一個帶事件特徵的詞組，"趁天黑"則是一個表示時間的成分。又如(21)

(21) 他趁這個機會打他一頓。

"這個機會"只是一個普通名詞，這裡的介詞片語是表示動作所利用的條件或機會。因此，針對電腦剖析語言的需求，如果介詞所引介的是時間名詞(組)，這個名詞所帶的時間特徵也會向上傳給介詞；如果介詞所引介的是 Event，我們則規定它們有特徵共存限制，隱含時間特徵，如下所示：

$[-N -V [+TEMP]] \& [+EVENT] \supset [+TIME]$

這麼一來，介詞組就帶有時間特徵。至於介詞和後續成分的共存限制，是爲了讓電腦能辨識時間成分的合法表示方式，如(22)，"自從"這個介詞本身就帶有從一個特定時間到以後的語意，所以它不能和"以前"之類的方位詞共同出現。

- (22) a. 自從他出國以後，就沒有消息了。  
b. \*自從他出國以前，就沒有消息了。

### 2.3 時間副詞

時間副詞在我們的日常言談中常常出現，但是它們多不表示"時"，而是表示"態" [陸儉明等 85]。例如：

表示過去時態：曾、曾經、已經、一度、早已、老早。

表示未來時態：早晚、終歸、即將、就要、快。

表示進行時態：正(在)。

這些訊息對於機器翻譯而言，也是很有幫助的，尤其中英翻譯所遇到的時態問題。目前爲止，我們尚未對時間副詞詳細分類，或許值得朝這方向去做。不過，時間副詞不具衍生性(*productivity*)，在辭典中是列得完的一類，而且出現的位置也很有規律，幾乎都是在動詞前面，要讓電腦辨識，不是一件難事。我們的作法是在詞項下給詞類及時間特徵即可。

### 3. 時間概念的分類與特徵

前面我們談電腦如何去判定時間詞組，那麼時間詞組出現的位置是不是固定呢？就我們分析文章後，發現大部分的時間詞組都出現在動詞之前。而一般人對於時間的概念大致可以分為三種：

一是時頻 (*frequency*)。表示動詞所表達的事件發生的頻率。它的形式特徵有“每”加時間量詞，如“每天”、“每年”；時間量詞的重疊貌，如“天天”、“年年”；還有頻率副詞，如“常常”、“經常”。時頻只能出現在動詞之前。如(23)

(23) 他每年去美國渡假。

另外，可能有人覺得在“他三天喝了三次水”中的“三天”有頻率的意味，實際上，它是受了“三次”的影響，“三天”應該還是指“喝了三次水”整個事件所花的時間長度。

二是指事件發生的時間，可以是時間座標上特定的一點，也可以是特定的一段時間。它的形式特徵分為：a. 有特指定詞或指示定詞[張等 88]修飾的時間詞(組)，如“這星期”、“去年”、“上個月”；b. 介詞時間片語，如“從昨天到今天”；c. 和方位詞構成的時間詞(組)，如“兩年後”；d. 無標記特指時間表示法，如“三點五十分”、“1983年”。以上的這些時間表示法，都是定指的，而且只能出現在動詞之前[註8]，如(24)

- (24) a. 我們今天上午十點鐘開會。  
b. \* 我們開會今天上午十點鐘。

三是指事件所花的時間或狀態持續的期間。它的標準形式是數詞或數量定詞加時間詞泛指時間的長度，如“三天”、“整年”；有時也在這些詞組後加上“之中”、“以內”強調時間的範圍。這一類可以出現在動詞前後，如(25)-(27)

- (25) 他三天沒吃飯了。  
(26) 他兩個月就完成了這個工作。  
(27) 他的手指腫了兩個月。

第二和第三類時間概念的區別，主要在於定指及非定指之分別。經由這些分類，我們得知在動詞後的时间詞組，是表示事件所花的時間或狀態持續的時間，定指的時間詞不能出現在這個位置。時間概念的分類，還是有助於電腦剖析時，理解句意以及辨別合法與不合法句子的能力。

#### 4. 結語

電腦剖析漢語，多以動詞的格框為重要剖析訊息[Chen et al. 88]，以找到動詞的必要論元為第一要務。本論文也是以這個原則為出發點，我們實際分析了八篇去年十二月號天下雜誌上的文章，發現時間詞組的問題不容忽視。首先要讓電腦剖析時，能判別那些是表示時間的成分，不致於和動詞的必要論元搞混。由本文第二節得知電腦基本上可以從三方面來判定時間成分：

1. 時間名詞(組)或時間副詞，直接由詞項上的時間特徵看出。
2. 帶事件特徵的詞組+時間方位詞，由特徵共存限制  
([+EVENT]&[TLOC] ⊃ [+TIME])得知。
3. 介詞+帶事件特徵的詞組，由特徵共存限制  
([-N -V [+TEMP]] & [+EVENT] ⊃ [+TIME])得知。

確定電腦能辨認漢語中的時間成分後，我們將三個時間概念粗淺地做一個界定，希望藉由不同的形式特徵以及出現的位置，進一步的幫助電腦理解語意或找出不合法的語句。當然，在這篇短文裡，想要完全掌握抽象而活潑的時間概念，似乎不簡單，例如“他每天跑三點半”這些習慣用法(*idiom*)，需要藉助經驗法則(*heuristics*)來處理，我們就不多作解釋。因此，本文完全是以電腦剖析語言的觀點，來研究整理時間成分，相信能解決大部分的問題。

☆☆本文之研究得中央研究院計算中心與工業技術研究院電子工業研究所合作之中文詞知識庫計畫(Chinese Knowledge Information Processing CKIP)之補助，並承陳克健、黃居仁兩位教授悉心指導以及詞庫小組各同仁閱後提出許多寶貴意見，本人由衷感謝！此外，謝謝楊淑吟、李惠君兩位小姐百忙中為我打出這篇論文。



### [附註]

1. 目前詞庫小組的分類系統中，只有氣象動詞這一類(如：“下雨”、“出太陽”)的主語(即必要論元)是時間詞。
2. 在詞庫小組的分類系統中，“前”是特指定詞、“三”是數詞定詞，這裡所指的定量式，可能是一個、或二個、或三個定詞和量詞構成定量式，至於定詞和定詞間的共存限制，見張等(1988:12)。
3. 雖然“今、明、昨、去、初、周”後面可接的時間量詞非常有限，但爲了語法分類上的一致性，我們仍將這些單詞放入定詞中。
4. “半”和“多”出現的位置有兩種，一是一般定詞所在的位置、即[量詞]，如：多年、半天；一是不同於其它定詞的位置、即[數詞定詞 量詞](時間名詞)]，如：三年多、三個半月。
5. 其中“世紀、年度、月、星期、禮拜、小時、學期”兼具名詞的特性。
6. “前、之前、後、之後、內、以內、之內、中、之中、間、之間、中間、當中、頭、裡、底”也是空間方位詞，例如：“桌子前”、“校園裡”。
7. 我們規定動詞組或句子一定帶有事件特徵。
8. 如果說定指的時間詞可以出現在動詞後面，那也只有右邊的這種情形：“他搭五點十分的(火車)”。“的”一定要保留，這裡牽涉到“的”的語法功能，本文暫不做進一步的探討。

### [參考書目]

1. 呂叔湘 (1984) 現代漢語八百詞 香港商務印書館。
2. 陸儉明、馬真 (1985) 現代漢語虛詞散論 北京大學出版社。
3. 陳克健，陳正佳，林隆基 (1986) “中文語句分析的研究--斷詞與構詞”，技術報告 TR-86-004，南港中央研究院資訊所。
4. 張麗麗等 (1988) 國語的詞類分析(修訂版) 技術報告 0002，南港中央研究院計算中心。

5. 魏文真、黃居仁 (1988) "中文的條件句"(待刊中)。
6. Chao, Yuen Ren (1968) A Grammar of Spoken Chinese. Berkeley and Los Angeles: University of California Press.
7. Chen, C.D., K.J. Chen & L.S. Lee (1986) "A Model for Lexical Analysis and Parsing of Chinese Sentences," *Proceedings of 1986 ICCCL, Singapore*, pp. 33-44.
8. Chen, K.J., L.L. Chang, C.R. Huang, C.C. Hsieh (1988) "A Classification of Chinese Verbs for Language Parsing," *CPCOL*, Toronto.
9. Gazdar, G., E. Klein, G. Pullum, I. Sag (1985) *Generalized Phrase Structure Grammar*, Cambridge: Harvard University Press.
10. Li, Charles N. & Sandra A. Thompson (1982) Mandarin Chinese: A Functional Reference Grammar. Berkeley and Los Angeles: University of California Press.
11. Paris, M.C. (1988) "Durational Complements and Verb Copying in Chinese," *Tsing Hua Journal of Chinese Studies, New Series XVIII*, No2, Hsinchu.
12. Sells, Peter (1985) *Lectures on Contemporary Syntactic Theories*, Stanford: CSLI.
13. Tai, James H.Y. (1984) "Verbs and Times in Chinese: Vendler's Four Categories," *Lexical Semantics*, pp.289-296, Chicago: Chicago Linguistic Society.