中央研究院歷史語言研究所集刊
第六十本，第一分（民國七十八年三月）
出版日期：民國七十九年三月

試論漢語的數學規範性質

黃 居 仁

本文介紹自然語言的數學規範理論，並依據此理論初步探討漢語之數學規範性質。利用規範語言之封閉性（closure），本文先證明漢語包孕句的現象超出尋常語言（regular language）。接著本文更利用漢語反覆是非問句的現變，證明漢語的規範性質超過免用語境（context free）語言。此一結果對自然語言之規範理論相當重要。其意義不僅在於再次提供自然語言爲超免用語境語言之實證，而且是第一個不藉同構轉換，直接以複寫形式（copying）證明的證明。

壹、前言：語言的數學規範性質

人類用語言以表達意見、溝通思想、傳承文化。至於語言的具體符號（口語的聲音與文字的形體）如何代表抽象的思維以及這些簡易有限的符號，如何反映具象世界的複雜現象與思維世界的微妙纖毫，則一直是哲學家、心理學家，甚至語言學家所苦思欲解的問題。在能夠解答這個有關語言與思考的大問題之前，西方語言學家所一直努力的方向則是先釐清語言這個表意系統內部的構造，其中包括了語句與其組成份子之間的關係。舉例而言，趙元任先生在他的 'Language and Symbolic Systems' (Chao 1968) 一書中就曾介紹了語言是傳遞資訊的符號系統這個觀念，並且略述了如何以數學方式計算語言承載的資訊及計算語言溝通的效率。近年來由於資訊科學的興起與電腦科技之影響，遂有數理語言學（Mathematical Linguistics，參見華爾 Wall 1972）之產生。語言既是傳遞資訊之符號系統，語言學家自可以將其抽象化，與其他符號系統相比擬，並研究其數學規範性質。這個研究方向，在純學術上可讓我們更清楚語言本身具有的數學特質以及語言和數學模式之間的關係；更可就規範性質界定「可能的語法」(Possible Grammars)，與「可能的語言」(Possible Natural Languages)，從而根據這些明確的性質評估斷定那些語法理論較爲可行。在實用方面，

科學家對處理規範化的符號系統（如數理邏輯，電碼，甚至電腦程式）已有相當良好的基礎與經驗；若能把語言與這些規範化系統比較並界定其數學性質，則許多已知的科學分析方式均可應用到語言分析上。具體而言，在運用電腦剖析（parsing）規範化的符號系統方面，已有相當明確有效的方法與策略；反而是在自然語言處理（Natural Language Processing）方面，由於對自然語言的數學規範性質不夠清楚，一直未能確定最有效的運算法（algorithm）。替語言的數學性質定位，將可直接影響電腦剖析語言模式及運算法之選定。

在進行討論之前，我們必須將語言一詞重新界定。在數理規範性質的探討中，「語言」一詞，如英文 language 一字，不但可指人類所用的自然語言（natural language），也可指人爲規定，電腦所用的程式語言（programming language）。換言之，在數理語言學與資訊科學的術語中，language 一字泛指所有可以傳遞訊息的符號系統[1]。在此一定義下，研究語言中那些成份可以成句（即那些由字符（Characters）所構成的字串（Strings）爲該語言中的合法語句），與這些成份之間的構成關係的學問稱之爲句法學（Syntax）。而研究成句的意義及其意義如何由各成份的意義組成導出，則稱之爲語意學（Semantics）。句法與語意同爲任何自然或程式語言所不可或缺的部份。本文所關切的是句法部份的數學性質。句法的表現形式，基本上是由一個起始符號 S（代表句子 sentences），一些非終端符號（non-terminal symbols, 如自然語言中的名詞組 NP，動詞組 VP 等），一些終端符號（terminal symbols, 即自然語言中的各個詞項，如中文「詩」、「書」），再加上由這些符號所構成的幾條改寫規律（re-writing rules, 如例〈1〉），所共同組成的。合法語句（grammatical sentence）的生成（generation），完全靠推演上述語法中幾個簡單步驟而導出。基本步驟是由起始符號 S 開始，利用所有可用的改寫規律，代換符號，直到整個字串（string）均由這個語言所認可的終端符號所組成爲止。這樣的一個字串即是該語言中的一個合法語句。相對而言，以此語法剖析語句的目的在判斷某一特定語句是否合語

1　爲求行文方便，本文中「語言」一詞的用法將採英文 language 一字的定義，即包括了自然語言，程式語言，及其他符號系統。至於中文「語言」一詞慣常所指的人類語言，則以「自然語言」一詞代替之。

法，故而剖析的策略便是以最簡潔的方法決定該語句是否可經上述步驟而由語法規律導出。

〈1〉 a. $S \rightarrow A\ B\ d$

b. $\alpha\ A\ \beta \rightarrow \alpha\ w\ \beta$

c. $B \rightarrow c\ d$

〔說明：S 爲起始符號

A B... 爲非終端符號

a b c... 爲終端符號

$\alpha\ \beta$ w 爲字串，其中w不可爲空字串（null string)〕

將語法如此規律化後我們便可研究語法的數學性質。杭士基（Chomsky 1963）根據對語法規律形式上的限制及其衍生能力的強弱將所有規範語法分成了第零型到第 3 型四類。第零型的語法，其規律無任何限制，故又稱爲「無限制改寫系統」（Unrestricted Rewriting System)。第 1 型的語法，其限制爲所有語法的規律必須爲〈1b〉的形式。也就是說，改寫規律必須同時規定其施用的前後語境。因此第 1 型語法又稱爲「需用語境語法」(Context-sensitive Grammar)[2]。

第 2 型語法則規定其語法規律必須如〈2〉的形式：

〈2〉 $A \rightarrow w$ （w不可爲空字串）

〈2〉規定的是第 2 型語法的形式必須是由一個非終端符號改寫成一個非空的字串（此字串中當然可包括終端符號與非終端符號）。亦即此型語法規律中只能規定某個項目能改寫成的項目而不能規定該規律施用時前後的環境。因而此型語法又名爲「免用語境語法」(Context-free Grammar)。

第 3 型語法規定其語法規律只有以下的形式：

2 陳克健教授指出 context-sensitive 之定義實容許語法規律不指定語境。即語境之存在並非必須。由此看來「需用語境」一詞並非最貼切的翻譯。但此譯名語言學界沿用多時，且似乎無與「免用語境」相對之其他貼切譯法，故本文仍採「需用語境」一詞。

〈3〉 a. $A \rightarrow \alpha B$ 或

b. $A \rightarrow \alpha$

（α 爲一非零字串，所有 α 的元素均爲終端符號）

〈3〉除了改寫時不能有前後語境的限制之外，更限制了改寫的結果最多只能含有一個非終端符號，而且該非終端符號只能出現在字串的最右邊[3]。第 3 型的語法又稱之爲「尋常語法」(Regular Grammar)或「有限狀態語法」(Finite State Grammar)。以上對於規範語言 (Formal Languages) 的分類，學者習稱爲「杭士基階層」(Chomsky Hierarchy)。

我們對語法規範性質的關切，基本上還是由於對這些語法所能產生的語言的關切。根據杭士基階層對語法的定義，這四型語法所能產生的語言也分成四型。尋常語法的限制最嚴，故所能產生的語言，其範圍最小。這些語言總稱第 3 型語言 (Type 3 Languages)。更通用的名稱是「尋常語言」(Regular Languages) 或「尋常集合」(Regular Sets)。第 2 型語法限制寬些，所產生的語言，其範圍包含並超過了尋常語言，除了稱爲第 2 型語言 (Type 2 Languages) 外，更普遍的名稱是「免用語境語言」(Context-free Languages)，因爲該型語法習稱爲免用語境語法。第 1 型語法又稱需用語境語法，其所生的語言包含並超出了免用語境語言。此型語言除了稱爲第 1 型語言 (Type 1 Languages) 外，也稱爲「需用語境語言」(Context-sensitive Languages)。至於第零型的語法則毫無限制，可以產生（規定終端與非終端符號後）所有的可能語言，當然也包含了以上提到的尋常語言，免用語境語言，與需用語境語言。此第零型語言 (Type O Languages) 又名爲「可遞廻列數集合」(Recursively

---

3　文中〈3〉的限制也可寫成 (i)

(i) a $A \rightarrow B\alpha$ 或

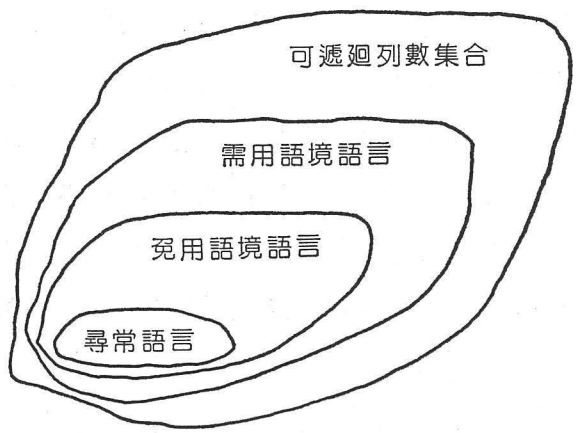
b $A \rightarrow \alpha$

如此一來，限制便成爲唯一的非終端符號只能出現在結果字串的最左端。這個限制下的語法所能產生的語言和〈3〉所限制的語法所能產生的語言是相等的。符合 (i) 限制的語法稱爲「左線性語法」(Left-linear Grammar)，而符合文中〈3〉限制的語法稱之爲「右線性語法」(Right-linear Grammar)。其名稱的來源是因爲兩種語法中的合法字串分別只能向左及向右擴充生長。

Enumerable Set)。此四型語言之間的包含關係可由〈4a〉及〈4b〉分別式示及圖示。

〈4〉 a. 尋常語言⊂免用語境語言⊂需用語境語言⊂可遞廻列數集合
（第 3 型） （第 2 型） （第 1 型） （第零型）

b.



很顯然的，根據以上定義，可遞廻列數集合包含了所有能以規範符號表達的語言，當然自然語言也包括在其中。換言之，第零型的語法定可產生所有自然語言，並且描述自然語言所有的正確結構。那麼，第零型的語法是否即提供了最理想的自然語言語法模式呢？答案是否定的。描述能力最強的語法，其所容許的可能語言結構形式也必然最多。也就是說，這一型語法不但可產生所有的自然語言，也同時容許了所有不可能出現在自然語言中的結構。問題是，我們寄望於語法的不僅是產生合法的自然語言語句，而且是正確有效地僅產生合法語句。若該型語法容許太多不合自然語言的現象產生，利用該語法的剖析運算法必得費時費力剔除這些現象。另外，語言學家也希望經由「可能的語法」(possible grammars) 來定義「可能的語言」(possible languages)。也就是說，他們希望找出恰能產生所有自然語言的語法出來；從而界定所有的自然語言可能出現的形式及語言之間可能的變異。以上兩定義可作評斷語法理論的絕對標準。假設第零型語法爲所有自然語言的語法模式，則不但引進許多不存在的語言形式，降低剖析、理解與習得的效率。更因爲包容了所有可能的規範語言，使得「可能的自然語言」的定義變成毫無意義。自然語言顯然與規範語言有別，且限制較多。以第零型語法等同於自然語言語法卻顯不出此項差異來。故學者尋求的是描述能力恰當而不是描述能力最強的語法。我們可試以一例說明描述能力過強的語法的缺

點。第零型的語法規律無任何限制，故其可出現的形式至爲繁複。上文中提到，語法的功能在判定某一特定的句子是否屬於該語法所產生的語言（即判定某一字串是否爲該語言的合法語句）。正因爲第零型語法容許產生所有的語言型式，任選一字串要判定該字串能否被特定的第零型語法產生這個問題，憑直覺而言，就好像在無限多的選擇項中要去碰一個相同的東西。試把這個問題簡化成在一袋數目無限多的色球中挑一白球。假設袋中根本無白球。因袋中球數無限，無法窮究，故而不管挑過的非白球有多少個，袋中隨時仍有無限多球待驗；故袋中無白球此一事實無法證明。設袋中有白球，但因袋中球數無限，無法確定白球會在第一次挑選，第一萬次挑選、或在第一千萬次挑選時出現。更因球數無限，剩下的球數不會減少，選中白球的機率亦不隨已挑選次數增多而增加。即無法斷定白球在某一有限時間內會出現。兩方相加，即袋中是否有白球這個問題無法在有限時間內得到解答。換句話說，比較第零型的語法和一個任意字串時，我們既無法在有限時間內斷定該字串爲合法語句，亦無法斷定該字串爲不合法語句；即以第零型語法判定任意字串的合法度這個問題爲不可解（undecidable）。我們當然不希望以此類型的規範語法作自然語言語法的模式。設若人類所用的語法眞是如此，小孩習得語言時應常有失敗（即學不會說話）的經驗，而人類聽話與說話時也會有頭腦空轉而無法解析或產生語句的現象，這和我們所知的語言現象是不相容的[4]。

資訊科學及數理語言學，對規範語言的研究中，另一個重要層面是找出和各型語法能力相當的「自動機」（automaton）。每個自動機都有一些固定的有限狀態、動作，及一些輔助的記憶裝置，可以直接轉換成電腦上的運算法。各型自動機能力之差異源自於其輔助記憶裝置所受之限制。有了這些嚴謹定義的自動機之後，我們便可用數學方式來測度各型語法判定語句的效率與範圍。在規範性質上，上一段所討論的第零型語法所能接受的語言範圍恰等於「杜林機器」（Turing Machines）所能產生的字串的範圍。杜林機器得名於英國數學家杜林（Alan M. Turing），數學上可模擬任何電腦上可執行的運算，是生成能力最強的運算模式。理論上有無限的記憶可供運

4　當然，這兒指的是正常狀況。人生病了，影響聽力或腦力，或環境噪音太大，影響聽力的情形都不在考慮之中。

用。問題是數學上可以證明判定任一字串是否爲某一特定杜林機器所接受這個問題爲不可解（undecidable）；亦即數學上不可能找出一個有效的運算法，在有限時間內找出這個問題的正確答案。對此，霍普克勞復興烏爾曼（Hopcroft and Ullman 1979）提供了相當完整的證明。這個結果和上段中直覺的描述相同。此一結果不但意味著第零型語法不是自然語言語法可行的規範模式；電腦處理語言時也不可用如此毫無限制的模式，否則將無法進行任何有效運算[5]。

一如第零型語言之對於杜林機器，尋常語言（即第 3 型語言）可由「狀態有限的自動機」(Finite State Automata) 產生；而免用語境語言（即第 2 型語言）則可由上述的狀態有限自動機另加一堆疊（Stack）產生[6]。數學上及資訊科學上對這兩種自動機均已有相當詳盡的研究，並已在電腦上設計了有效的剖析運算法。如能把自然語言相對於這兩型語言作一定位，不但有簡潔且數學性質清楚的規範模式可用；在自然語言處理上更有既成有效的運算法借用。再者，如果我們能把自然語言的規範性質確切定位，我們可以用之與語法理論的規範性質相比。良好的語法理論其規範性質必須與自然語言的規範性質相當。即若自然語言恰等於第 2 型語言，正確的語法理論也必須恰是第 2 型的語法。若語法過簡易，則自然語言中有些語句無法產生，當然不是良好的理論。若語法過於繁複，則不但會蔓生（over-generate）；亦會使判定某一語句是否合語法的問題過於複雜；無法在有效時間內解答。因之，語言與語法的規範性質

5 事實上，十數年來新語法理論如概化詞組結構語法（GPSG）與詞彙功能語法（LFG）之棄變形 (transformation)，以及管理約制理論（GB）之對變形律加以設限簡化，受佩德斯與瑞奇（Peters and Ritchie 1973）的影響甚大。兩氏之文在數學上證明當時衆人接受的變形語法的生成能力相當於杜林機器。也就是說對於任何句子是否合於某個變形語法規定這個問題在數學上根本不可解。除非在接受此一事實後又能解釋人類習得及理解語言的現象，否則，變形語法絕非可行的自然語言語法模式，已無可置疑。

6 狀態有限的自動機僅有固定狀態而無記憶裝置，是最原始的計算模式。堆疊是一種記憶裝置。我們可以將其想像成自助餐廳中存放碗碟的架子。每個堆疊可以容納一定數目的碟子；但每碟置入時均恰在舊碟的上方，因此最先置入的碟子在最下層，緊接而來的碟子依序置放其上，而最後置入的碟子在最上方。取碟子時只能由最上方取用，因此只能取得最新的碟子；而先置入的碟子則等所有其他碟子都取得之後，最後方能取到。這和人類的記憶有一點類似，即最近接觸的事物愈易記起；而「記憶深處」的資料得稍費周章方能取得。另外，第 1 型語言（需用語境）則可由線性受限（linear-bound）之自動機模擬。

研究，對於語言學家及心理學家之尋求正確描述人類語法知識，與電腦學家與語言學家之尋求有效剖析自然語言之運算法均爲不可或缺之基礎研究。

貳、目標與方法

對於漢語數理邏輯特性之研究，趙元任先生早於一九五五年及一九五九年分別發表了兩篇文章，收錄在他一九七六年出版的論文集中 (Chao 1976a & 1976b)。此兩篇文章探討了各種邏輯關係在國語中的表達方式。至於眞正討論漢語數學規範性質的文章則不曾見。本文的主要目標之一乃在於初步探索界定漢語的一些數學規範性質，以作未來漢語規範語言學及電腦剖析中文研究之基礎。

其次，有關所有自然語言規範性質之探討，如自然語言到底是包含於免用語境語言或超出免用語境語言而包含於需用語境語言，一直是學術界爭論極多的問題。蓋志達 (Gazdar 1981) 指出了若自然語言包含於免用語境語言在理論上的好處在於能解釋人類迅速剖析及習得語言的現象。蒲倫與蓋志達 (Pullum and Gazdar 1982) 則以相當嚴謹的數學及語法論證駁斥了所有變形語法學者所提出的，認爲自然語言超過免用語境語言的論證。卡司 (Kac 1987) 及卡司等人 (Kac et al. 1987) 卻又在最近同樣的利用英文 'respectively' 的並列結構加上更複雜的數學步驟欲求翻案[7]。其實，目前爲止唯一爲所有學者所接受，自然語言超出免用語境語言的證明是由施別 (Shieber 1985) 所提出。該文是利用瑞士德文 (Swiss German) 關係子句中交替序列對應 (cross-serial dependencies) 的現象，轉換成同構 (homomorphic) 的規範語言，

7　英文 "respectively" 的用法和國語「分別」一詞類似。雖然國語「分別」一詞似乎容許比英文 "respectively" 一詞更爲自由的解釋。

(i) 張生、王生、李生分別選了歷史、語言學、考古學。

(i) 句中之個人與他們所選讀的學科之間的關係可由連線表示。以符號簡化，此句中的關係近於 abcabc 這樣的字串。即將整個字串剖爲兩半，前半的第一字需對應於後半的第一字，第二字對應於第二字。此種交替字列相應的語言爲免用語境語法所不能處理，下文中將有討論。問題是「分別」(與英文 "respectively") 並未規定句子一定要有 (i) 的形式。

(ii) 這三個學生分別選了歷史、語言學、考古學。

(ii) 句亦是合法的中文句。雖然其主語只有一個名詞組，不能產生交替序列相應的現象。因此 (i) 的現象不是語法的現象，不能用以證明國語的語法規範性質。蒲倫與蓋志達(Pullum and Gazdar 1982) 以英文提出了類似的論證。

然後證明該語言爲需用語境語言，因而證明該語言超出免用語境語言的範圍[8]。如果自然語言僅有此一特例超過免用語境語言，則其象徵性可能超過實際研究規範語法時的重要性。也就是說，蓋志達等人（Gazdar et al. 1985）以免用語境語法分析自然語言的理論仍有其可行性。同時，瑞士德文的例子是句子前後半段有固定關係，不是前後半段字皆相同。因此，瑞菁斯基（Radzinski 1987）也指出如果某個自然語言中有不需轉換成同構關係即可證明是超免用語境語言的現象，對於確切爲所有自然語言的規範性質定位，將是更強而有力的證據。筆者認爲漢語中有數個這樣的現象。因此，本文除了探討漢語的規範性質外，也將爲自然語言的性質超出免語境語言這個事實提出明確的證據。

證明某個語言的型屬最直接了當的辦法是構建一個僅能衍生該語言的語法，然後在數學上驗明該語法的規範性質與型屬。可惜在探討自然語言時此一方法並不可行。由於自然語言的現象過於龐雜，至今語言學家的研究尚未寫出任一語言完整的語法出來，更遑論深究其規範性質了。至於取語法的片段來研究其規範性質更不可行；片段語法的規範性質不見得等於整個語法的規範性質。如需用語境語法的片段可以是免用語境語法，亦可以是尋常語法；故鑑定語法片段的規範性對決定整個語法的規範性並無太大幫助。

對於證明自然語言規範性最有用的數學性質是這幾型語言的封閉性（closure），以及與某些已知其類型的人造語言相比較。尋常語言、免用語境語言及需用語境語言對於代換（substitution）這個運算均有封閉性。也就是說我們可以對某個尋常語言的

---

8 施別（Shieber 1985）所舉的瑞士德文例子與底下所舉的荷蘭文例子非常相近（兩者均爲西日耳曼語的分支）。但荷蘭文因有其他可能語序（word order），故蒲倫與蓋志達（Pullum and Gazdar 1982）及布瑞斯冷等人（Bresnan et al. 1982）均曾指出該語言事實上可以免用語境語法產生（儘管所生的結構可能不正確）。瑞士德文和荷蘭文的差別在於某些特殊的格位。至於這些差別對規範性質的影響請參看所引的三篇文章。以下例子中所顯示的交替序列相應現象則是兩個語言所共有。

(i) ...dat Jan Piet de Kindren zag helpen zwemmen（荷蘭文）

〔that〕	尙	彼德	孩子們	看見	幫助	游泳
	NP1	NP2	NP3	V1	V2	V3

「（……）尙看見彼德幫助孩子們游泳。」

NP1、NP2、NP3 分別是 V1、V2、V3 的主語，其關係可簡化成上述的 abcabc。

幾個字符或一個長度有限的字串進行代換，而其所產生語言必然依舊是尋常語言；餘依此類推。由於代換的定義可把一個字符換成另一個字符，把一個字串中的數個字符換成一個字符，甚至把這些字符換成空字串 (null string)。在探討自然語言時最常用的技巧便是把句子與規範性質不相干的部份代換成空字串，以利於凸顯我們所要探究的部份。另一個更重要的封閉性是此三型語言與尋常語言的交集均有封閉性。也就是說尋常語言和尋常語言的交集是尋常語言，免用語境語言和尋常語言的交集是免用語境語言，而需用語境語言和尋常語言的交集則是需用語境語言[9]。大部份對自然語言規範性的反面證法均利用此一封閉性。即先假設某語言 L 爲第 n 型語言，然後再構建一尋常語言 R 使得 R 與 L 的交集非爲第 n 型語言。由於第 n 型語言與尋常語言之交集必需亦爲第 n 型語言。此結果顯示假設錯誤，L 的規範性必須超過第 n 型語言。以上兩個封閉性的數學證明均見於霍普克勞復與烏爾曼 (Hopcroft and Ullman 1979) 'Introduction to Automata Theory, Languages and Computation' 一書中。

叁、漢語的數學規範性質

叁之一　漢語並非「尋常語言」

要證明漢語不屬於尋常語言，我們必須利用尋常語言與尋常語言之交集的封閉性。換句話說，因爲尋常語言對交集具有封閉性，因而兩個不同的尋常語言的交集仍是尋常語言。當某語言甲非常複雜，其數學性質又不明確時，欲證明其並非尋常語言，最簡單的方法之一便是拿甲語言和一個已知的尋常語言乙交集，然後再證明所得的結果丙並非尋常語言。由於我們已知乙是尋常語言，根據尋常語言的封閉性，設若甲爲尋常語言則其交集丙亦必須爲尋常語言。已知此推論不眞，故可推知其假設爲

9　此處所談各型語言交集的封閉性與圖〈4a〉所示各型語言之間的包含關係並不衝突。該圖所顯示的乃所有各型語言所成的集合之間的關係。例如需用語境語言所成的集合包含了免用語境語言。意即每個免用語境語言亦是需用語境語言；每個免用語境語言均可由相當的需用語境語法產生。此處所提的則是單一語言之間的交集。同一類型各個單一語言的定義彼此之間可大相逕庭，其交集也就不見得保存原來的規範性質，單一免用語境語言和其他免用語境語言的交集即無封閉性。比如 $a^ib^ic^j$ 與 $a^ib^jc^j$ 均爲免用語境語言，但此兩語言之交集則爲 $a^nb^nc^n$，爲一熟知之需用語境語言。

僞。結論爲甲並非尋常語言。

對此一規範性質時所涉及的國語結構爲包孕(central embedding)的關係子句。國語的名詞組可以由關係子句構成，如〈5〉

〈5〉 認識胡適之先生的人

可是這整個由關係子句所構成的名詞組亦可包孕在另一個關係子句中，形成另一個由更複雜的關係子句所修飾的名詞組。國語中可能出現〈6b〉那樣的名詞子句，當然也可以有〈6a〉那樣的名詞子句。

〈6〉 a 認識〔np 認識胡適之先生的人〕的人

b 訪問〔np 認識胡適之先生的人〕的人

〈6a〉這樣的關係子句可以無限擴充下去，雖然聽者或讀者理解這樣的結構稍有困難，只要「認識」一語重複的次數和「的人」一語重複的次數相同，都可算是合法的中文詞語。即理解〈6〉這樣的句子受限於人類剖析語句的運算法與策略，卻不受限於語法。

根據定義，尋常語言中所有的語句都能以「尋常表現式」(regular expressions)，代表，而所謂尋常表現式，則是除了按順序列字串外，僅能利用柯寧星號（Kleene star)「＊」，來標示某個字符可重複零至無限多次。模仿上面提到的國語關係子句，我們可以下面的尋常表現式定義一個假想語言。

〈7〉 Ro＝{(認識)* 認識胡適之先生的人 (的人)* 很多}

上式所產生的所有字串集合成一個尋常語言，這個語言 Ro 中的合法句子必須有「認識胡適之先生的人」這一個字串出現；而在這字串之前，「認識」這個詞語可以不出現（即柯寧星號所容許的所謂重複零次），可以出現一次，也可以重複出現任何次數；相似的，在這個字串之後，「的人」，這個詞語可以不出現，可以出現一次，也可以重複出現任何次數。拿〈7〉所表達的語言與國語交集（即取這個語言中產生，而在國

語中也合法的句子）。我們發覺，國語中接受的句子必須是前後兩個語詞重複的次數一致。例如：選擇重複次數爲O時我們得到〈8a〉，爲2時得到〈8b〉。

〈8〉 a　認識胡適之先生的人很多

b　認識〔認識〔認識胡適之先生的人〕的人〕的人很多。

簡言之，國語和〈7〉所產生的尋常語言交集的結果可由下列的數學方式定義。

〈9〉 $L_0=\{(\text{認識})^n\ \text{認識胡適之先生的人}\ (\text{的人})^n\ \text{很多}\mid 0\leq n\}$

在上式中n是一個大於或等於0的變數，可是由於這個變數在同一個式子中出現，故其值不能改變。即（認識）和（的人）重複的次數必須是相同的。可是〈9〉並非尋常表現式，因尋常表現式只能有柯寧星號，而不能指定某一字符出現的順序。在數學上，上述表達式可經由代換步驟簡化成 a^nb^n。a^nb^n 的語言不能由「狀態有限自動機」(Finite State Automata) 產生，故不是尋常語言。由上得知，國語與一個已知尋常語言交集的結果不是尋常語言。故國語的數學規範性質超過尋常語言，而並非尋常語言[10]。漢語其他方言如閩南語、粵語亦有類似的關係子句結構，依同理可證其並非尋常語言。

漢語的規範性超過尋常語言（或杭士基階層中的第3型語言），此一結果並不出人意料之外。往更複雜的一層看，我們接著要探討的是漢語是否是免用語境語言（第

10　事實上，國語的關係子句結構可用兩條簡單的詞組結構律 (phrase structure rule) 表達出來。

(i) NP→VP 的 N

(ii) VP→V NP

由於規律 (ii) 右邊的 NP 可重複套用此一組規律，故可產生無限次的重複。但因規律用一次即產生一次 VP，故動詞重複的次數和名詞重複的次數一定相同。此一規律爲免用語境的語法規律。至於代換簡化之後的形式 a^nb^n 則可由下列簡單的免用語境語語法產生。請注意此語法雖簡單，却無法轉換成右線性語法或左線性語法，故產生的語言特性超出尋常語言的範圍。

(iii) S→aSb

(iv) S→ϕ

2 型語言）。

參之二　漢語並非免用語境語言

本節中將探討三個漢語的語法結構，每個結構均以國語舉例，並證明這三個結構的特性都不是免用語境語法所能描述，因而證明漢語的規範性質超過免用語境語言，據此推論，自然語言整體而言亦是如此。

在此節中的證明需用到一個熟知的需用語境語言（即其規範性質超過免用語境語言）。這個語言在文獻中通稱爲 ww 語言，或稱複寫語言（copying language）。即是由兩個相同的小每串結合成的語言。我們可以用下標（subscript）將其更明確地表達成〈10〉

〈10〉　$x_1, x_2 \cdots\cdots x_i \cdot x_1 x_2 \cdots\cdots x_i$

上文中提到免用語境語法可用一個帶單一堆疊的狀態有限自動機來模擬。故 $a^n b^n$ 這樣的語言和鏡像語言（mirror language, ww^{-1}，即每字串的後半段恰似前半段倒過來，如 abba, abcddcba 之類）均可由此型語法產生。此一系統可以加以簡化描述，狀態有限的自動機由左至右處理字串，每進一步即讀入字串中的一個字母，但不留下任何記錄。堆疊是唯一儲存訊息的地方，但因只有一個堆疊，故只能存一種訊息。對於 $a^n b^n$ 這個語言，所儲的訊息是 a 的個數，自動機每讀一個字符 a，堆疊即放入一片新碟，處理了 n 個 a 便堆了 n 個碟。自動機讀到 b 時，每讀入一個 b 字母即由現有堆疊最上層取出一碟；讀入了 n 個 b，便取出 n 個碟。由於系統規定字串處理完畢時堆疊必須恢復原始狀態（即爲空堆疊）；而且該取碟時不可無碟可取。因此前半段儲入碟的數必等於後半段取出的碟的個數，即此系統所容許的字串 a 與 b 數目一定相同。對於鏡像語言，則是在處理前半段字串時每讀入一字符即儲入一相對的新碟，即讀入 a 時即放入 a 碟，讀入 b 時即儲入 b 碟，亦即堆碟只儲身份關係，而不認其順序。到了後半段則是讀入一字符即取出一相對的碟，即讀入 a 時取 a 碟。由於堆疊後進先出（last in, first out）的特性，後半段字串正好倒了過來，成爲鏡像語言。至於 ww 語言，由於前後兩半段順序相同，亦即語法不但要記得字串中每個字符的值，

亦要知道順序關係；超過本系統的能力，故無法處理[11]。

叁之二・一　國語的反覆問句

國語中是非問句的形式之一是加入一個否定詞「不」或「沒」之後再在其後重複整個謂語。此類問句在文獻中習稱A非A (A-not-A) 或V非V (V-not-V) 問句 (Chao 1968, Huang 1982, Tang 1986)，雖然較口語化的語法會將部份相同的詞語省略，〈11〉這樣的句子略嫌累贅，卻完全合乎國語語法。

〈11〉　你看過杭士基的書沒看過杭士基的書？

黃正德 (Huang 1982) 與賀克司馬 (Hoeksema 1987) 均曾提起國語這個結構似乎不是免用語境語法可處理的，瑞青斯基 (Radzinski 1987) 更嘗試提出了數學規範證明。本文將更仔細嚴謹證明此一現象的超免用語境性 (Supra-context-freeness)。欲探索此一構造的規範性質，我們必須闡明在否定詞前後的兩個成份的對等關係純粹由語法限制。基本上，反覆是非問句前後兩成份並非是詞類相同即可，否則這類句子由簡單的詞組結構律即可導出。〈12〉句中否定詞前後均爲 VP，在語法上地位相同，〈12〉卻是不合法的語句[12]。

11　類似系統加上第二個堆疊即可處理，但雙堆疊系統相應的語法已不是免用語境語法，是需用境語法了。

12　請注意語言學家傳統上以句前的星號表示該句不合語法。千萬不可與前文中符號右上角標示重複的柯寧星號混淆。Carl Pollard 指出一般人在討論自然語言的規範性質時，往往忽略了語法的弱衍生能力 (weak generative capacity)。也就是說，某個自然語言即使有某些局部的需用語境現象；這些現象可能只是限制某一類句子而非限制整個語言。舉例說明，甲語言可能以複寫 (ww) 語言的形式表達問句。可是並不見得可以證明甲語言即是超免用語境。在證明過程中，我們也可能發現所定義的尋常語言和甲語言的交集只是個免用語境語言。因爲甲語言中可能有和上述複寫句型互補的句型；使得單純的免用語境語法即可產生該語言中所有合法語句，包括了 ww 語句及和其互補的句型，而不需借助需用語境規律。換句話說，免用語境語法弱衍生 (weakly generate) 該語言。在這情況下，我們通常仍視該語言爲免用語境語言。蒲倫與蓋志達 (Pullum and Gazdar 1982) 即指出了不少前人在證明超免用語境語言時得到的錯誤結論均由於未能排除上項可能性。
瑞菁斯基 (Radzinski 1987) 的證明中即未提出所有未合複寫語言要求的句子便不是中文的合法句。我們很容易揣測前後不相等的 V-not-V 問句是否可當選擇問句，如 (i)。

〈12〉 ＊你看過杭士基的書沒聽過貝多芬的音樂?

其次，我們要證明此類問句否定詞前後成份相等的限制並非是語意的限制。也就是說，我們必須確定反覆是非問句前後兩成份相等的限制是語法規範的限制而不是知識或邏輯的不合理。〈13a〉及〈13b〉顯示前後兩成份即使語意相當，用詞仍不可相異。

〈13〉 a ＊你看過杭士基的書沒讀過杭士基的書？

b ＊你看過杭士基的書沒看過瓊士基的書？

在〈13a〉中「讀」與「看」兩個動詞在此語境中幾乎是同義詞，但仍不可同時出現在此類是非問句中的並列兩項。〈13b〉更明確顯示語意條件不能決定反覆是非問句的形式。Chomsky 一字在中文中語音更精確的譯法爲「瓊士基」。可是即使說話者知道「杭士基」與「瓊士基」兩詞的指涉（reference）同爲 Chomsky 其人，〈13b〉仍爲不合語法的句子。也就是說〈13b〉可並列的兩個動詞組儘管長度相同，語音近於相等，而且語意完全相同；仍未具有合法國語反覆是非問句的充分條件。

最後，我們可由語言學理論的觀點來討論反覆是非問句的合法度非由構詞（morphology）決定。「反覆」（reduplication）爲極常見的構詞現象。然而語言學家對構詞在語法理論中的地位看法雖不盡相同；或主張自主的（autonomous）構詞學，即認爲構詞部份的規律自成一單元（module），而不與語法等其他單元發生關係；或主張構詞的規律應在語法規律之前；或主張構詞規律與語法規律平行，形式亦相同。所有語言學家均同意的卻是構詞規律中不可使用語法規律供給的訊息。可是，很顯然的，國語反覆問句結構中反覆的部份必須是完整的詞組。但詞組的構成成份則完全由語法決定。反覆問句的反覆現象涉及語法訊息，故不可能是構詞的重複現象[13]。

(續)(i) 你看過杭士基的書還是聽過貝多芬的音樂。

(ii) (=〈12〉)* 你看過杭士基的書沒聽過貝多芬的音樂。

(i) 句顯示中文選擇問句中用連接詞「還是」。(i) 句爲合法句；(ii) 則否。反覆是非問句只能用「沒」或「不」；而此類句子前後兩成份不是逐字相等時即爲不合法句，也無選擇問句的解釋。故此我們可證明國語的規範性質實在超過免用語境語言，無法以免用語境語法弱衍生。

13 筆者認爲即使反覆是非問句結構被重新歸類爲構詞現象，文中利用此結構對漢語規範性質的探討仍有效。因爲構詞其實也是描述語言規範性質規律的一部份。蓋志達與蒲倫（Gazdar and Pullum 1985）討論自然語言規範性質時也涉及構詞。

以上的討論確定了國語反覆是非問句之合法度既非取決於語意亦非取決於構詞，亦即此一結構之產生完全取決於語法。因此底下對此一結構規範性質之探究可直接解釋爲國語語法規範性質之表現。

首先，我們必須證明此類句子，除了受人類發音能力與記憶之限制外；長度可無限延伸。此一證明的主要原因在長度有限的句子雖具有複雜的規範性質，通常仍可以較簡單的語法或甚以列舉的形式產生，故不成爲該語言規範性質的證明。國語，和其他語言相同，可在名詞前加個數不限的修飾語，故其長度可無限延伸。而國語反覆問句中當然也可加名詞，故其長度也可無限延伸。

探討國語規範性質之第一步爲定義一個適當的尋常語言，俾與國語交集時產生有意義的結果。試定義如〈14〉。

〈14〉 R_1＝{你閱讀有益 α 的書不閱讀有益 β 的書｜其中 $\alpha, \beta \in$ {又有趣，又易讀*}}

R_1 的定義是說該語言所有的「句子」均由「你閱讀有益」開頭，由「的書」結尾；而中間則由兩個小字串夾著「的書不閱讀有益」而組成；至於這兩個小字串則規定是由選擇「又有趣」，「又易讀」這兩個詞作任意排列組合而成，柯寧星號規定任一個詞均可重複出現由零至無限大的任意次數。此一語言爲一尋常語言。

以 R_1 和國語交集，所得的語言是由既是國語的合法句，又屬於定義〈14〉描述的所有句子所組成。〈14〉定義的語言 R_1 恰和國語反覆是非問句重疊。而國語反覆是非問句中否定詞前後兩個詞組要逐字相等。故R與國語交集的結果可由〈15〉表示。

〈15〉 L_1＝{你閱讀有益 α 的書不閱讀有益 α 的書｜其中 $\alpha \in$ {又有趣，又易讀*}}

請注意 L_1 和 R_1，不同的唯一一處在於 L_1 中規定前後兩個字串變數需完全相同（即同以 α 一個變數代表）。也就是說 L_1 產生的句子必須是如以下的形式：

〈16〉 a. 你閱讀有益又有趣又易讀的書不閱讀有益又有趣又易讀的書

b. 你閱讀有益又易讀又有趣的書不閱讀有益又易讀又有趣的書

利用代換，我們可把 L_1 中以文字寫明的部份消去（因規範語言中以任意符號，即使是空號 ε 代替其他字母均不改其規範性質）。如此 L_1 可簡化如下：

〈17〉 $L_1' = \{\alpha\alpha \mid 其中\ \alpha \in \{\overset{*}{又有趣，又易讀}\}\}$

L_1' 很顯然是由兩相同字串組成的語言，即上文提過的複寫語言。此一語言超過免用語境語言，爲需用語境語言[14]。國語和一個尋常語言的交集不屬於免用語境語言。利用免用語境言和尋常語言交集的封閉性，我們可以得到結論。國語並非是免用語境語言，國語的語法也必須超出免用語境語法。

叁之二・二　其他前後重複之語法結構

國語中另有其他結構涉及在否定詞前後反覆詞序完全相同的詞組。其中「不管」這個動詞與上節討論的反覆是非詞句有直接的關係。「不管」一詞後面需接一賓語。而這個賓語必須是個名詞，如〈18a〉，或是一個疑問句，包括了反覆問句，如〈18b〉。事實上，完全不省略的反覆是非問句在當「不管」的句賓語時聽起來比獨立出現時更易接受。

〈18〉 a. 他才不管你的死活。

b. 他才不管你喜歡吃香蕉蘋果不喜歡吃香蕉蘋果。

我們可以證明「不管」的句賓語受到的語法限制和獨立出現的反覆是非問句相同。比如說，〈18b〉中的句賓語是一個並列的名詞組。國語名詞的並列結構次序調動並不影響其語意與語法。在本例中，「香蕉蘋果」與「蘋果香蕉」同爲名詞組，意義也相同。可是若在「不管」的句賓語中將其中之一代換，句子則變成不合語法，如〈19〉。

〈19〉 ＊他才不管你喜歡吃香蕉蘋果不喜歡吃蘋果香蕉

14 ww語言不是免用語境語言的完整數學證明請參閱霍普克勞復與烏爾曼(Hopcroft and Ullman 1979)，136 頁。

利用此一現象證明國語規範性質時可定義尋常語言如下：

〈20〉 $R_2 = \{$他才不管你喜歡吃 α 不喜歡吃 β $\mid \alpha, \beta \in \{$香蕉，蘋果$\}^*\}$

即 R_2 中的句子包含了兩個由任意數目的「香蕉」與「蘋果」作任意排列組合而成的名詞組。此語言與國語的交集爲 L_2。

〈21〉 $L_2 =$ 他才不管你喜歡吃 α 不喜歡吃 α $\mid \alpha \in \{$香蕉，蘋果$\}^*\}$

L_2 只比 R_2 多了一個限制，即兩個可代換的字串必須在成份及排列順序上完全相同，把 L_2 代換簡化成同構的 L_2'，同時把「香蕉」以 a 代替，「蘋果」以 b 代替。我們又得到一個標準的複寫語言。

〈22〉 $L_2' = \{\alpha\alpha \mid \alpha \in \{a, b\}^*\}$

由於 L_2'，並非免用語境語言，因而 L_2 也不是。利用免用語境語言的封閉性，國語非免用語境語言再度得證。

另一個類似的國語結構爲表示某一條件不存在之「無條件句」，如〈23〉[15]。

〈23〉 雙黃線不雙黃線，他超車可是照超不誤。

基本上，任何名組均可出現在「不」字的前後[16]。如〈24〉

---

15 語法學家或會認爲此類句子是「不管」、「不論」、「無論」等詞的省略。問題是可省略的詞並非僅有一個，故無法恢復（recover）被省略的成份。更重要的，如（i）這樣的句子加上上述任一個詞似乎都不甚好。筆者仍遵循表面語句爲準（surface-based）的重要假設；主張表面不同的語句語法結構亦應分開處理。
(i) 玉皇大帝不玉皇大帝，拜拜不該如此浪費。

16 (i) 下雨不下雨，孩子們得上學。
如 (i) 這樣的句子中「下雨」可視爲句子或名物化的名詞組。本文只考慮名詞組的句子，因爲非名詞組的重複條件不盡相同。如（ii）中可確定重複單位爲動詞組，此時前後兩動詞組不見得要完全相同。而且此時語意語法也似乎相異。即第一個動詞組似乎是主題，而不再是無條件句。把討論限制於「名詞——不——名詞」可避免混淆不同的結構。
(ii) a 讀書不讀書，這麼晚了不許看電視。
b 讀書不好好讀書，這麼晚了不許看電視。

〈24〉 一千萬不一千萬，學術良心可是不能出售的。

相對於〈24〉，我們可定義尋常語言 R_3。

〈25〉 $R_3=\{$一千 α 萬不一千 β 萬，學術良心可是不能出售的 $\mid a, b\in\{$一千，一萬$\}^*\}$

R_3 的定義會產生如〈26〉的句子。

〈26〉 a. 一千一萬萬不一千一萬萬，學術良心可是不能出售的。
b. 一千一千萬不一千一千萬，學術良心可是不能出售的。
c. 一千一萬萬不一千萬，學術良心可是不能出售的。
d. 一千一千萬不一千一萬萬，學術良心可是不能出售的。

很顯然的，〈26a〉和〈26b〉是好的國語句子，而〈26c〉和〈26d〉則不是。即「不」前後的名詞組要逐字相等。故 R_3 和國語的交集是 L_3。

〈27〉 $L_3=$一千 α 萬不一千 α 萬，學術良心是不能出售的 $\mid \alpha\in\{$一千，一萬$\}^*\}$

同樣的，我們證明了此類無條件句的名詞組雖可無限擴展，但前後兩個名詞組仍需字字對等，不可有任何差池。將 L_3 除了 α 以外的字代換以空字串，將「一千」代以 a，「一萬」代以 b，所得和 L_3 同構的 L_3' 事實上即是我們所熟悉的典型非免用語境語言。

〈28〉 $L_3'=\{\alpha\alpha \mid \alpha\in\{a, b\}^*\}$

因此，利用無條件句，我們也證明了國語和尋常語言 R_3 的交集不是免用語境語言。由免用語境語言的封閉性，再度推論出國語的規範性質超過免用語境語言。

叁之二・三　漢語中的對仗詞「對」

對仗是中國文學特有的現象。國語中的動詞「對」則專指此對仗的現象。若以最常見的一對春聯爲例，可有底下的句子。

〈29〉　「天增歲月人增壽」對「春滿乾坤福滿門」。

本文之重點不在討論對仗之規律而在討論「對」這個動詞的語法性質。簡單說來，「對」這個動詞需要一個主語與一個賓語。但是，其主語與賓語之間必須有一特殊關係，即兩者的長度必須相等，而且兩者內部的語法成份結構必須相同。此條件的第二部份可用杜甫月夜憶舍弟詩中的一聯來說明。

〈30〉　「露從今夜白」對「月是故鄉明」。

我們知道此兩句詩分別是「從今夜白露」和「是故鄉明月」倒裝的結果；把「白露」與「明月」這兩個雙音節詞拆了。這是詩中特殊的現象，口語句法中不能容許這樣的倒裝，可是，儘管此兩句有如此特殊的結構。〈30〉句仍爲合法句。即主語、賓語仍可分析成「名詞——動詞，名詞（雙音節）——形容詞」的對等字串而被接受。反之，〈31〉中的賓語語意雖相同，但因其成份結構不同，即不被接受[17]。

〈31〉　＊「露從今夜白」對「是故鄉明月」。

「對」字主語、賓語長度均無限制。雖然實際上五言、七言最常見，然對仗本身並無具體的長度限制。輓聯的長度使用自由，而且通常不只七言，因此，我們觀察到的現

17　史語所同仁指出此聯多數人接受的解釋是僅作字面意義解；而非筆者所採用，用仇注「時逢白露節」衍化而出的倒裝句解。魏培泉先生指出另一首杜詩有將名詞組拆散之現象。此聯出自秋興八首之八。

(i)　紅豆（或作香稻）啄餘鸚鵡粒，碧梧棲老鳳凰枝。

此聯似應解爲「鸚鵡啄餘紅豆粒，鳳凰棲老碧梧枝。」唯解杜詩者亦有人，如蒲起龍，將「鸚鵡粒」，「鳳凰枝」均釋爲名詞組。模稜矇矓本是詩句美處之一。我們似可不必在此追究杜工部原意。不論何解爲正確，詩聯逐字詞類相等的限制均是一致的。本文立論不受影響。

象是「對」字要求其賓語與主語爲長度相等，且語法上的類別與先後順序一一相應。將此一現象作同構轉換之後，涉及「對」的句子可簡化成 ww 的形式。（因其字串前半段與後半段詞類依序相應，故可將相同詞類的成份代以相同符號以彰顯其對應關係）。再仿上節中的證明方式，定義一適當的尋常語言與國語相交集，使得重疊的部份恰是動詞「對」所管轄的句子。交集的結果爲一非免用語境語言，由封閉性得證國語的規範性超過免用語境語言。

以「對」字涉及的對仗現象證明國語的規範性質，其實會遭遇和註七中所討論有關「分別」這個結構一樣的顧慮。即「對」字對其主、賓語的相應成份結構限制到底是語意上的呢還是語法上的呢？若是語意上的限制則無涉於國語的規範性質。由於上節中已提供了足夠的證據證明國語非爲免用語境語言；關於「對」字結構限制的語法性，本文暫不下定論。不過，筆者仍需指出，即使「對」字的限制是語意上的，仍要涉及語法知識。也就是說，該限制包括了相對成份的語法上的類別需相等[18]。

肆、結　　語

詞彙功能語法足以產生某些需用語境語言，此一事實開普南與布瑞斯冷 (Kaplan and Bresnan 1982) 曾加以證明。該文描述了詞彙功能語法如何產生 ww 語言。其他許多新的語法理論也都有處理某些需用語境語言的能力。賈恤 (Joshi 1985) 提出「略需語境語法」(Mildly Context-sensitive Grammar) 一詞。並指出自然語言超出免用語境語言的規範性質有限，故不需要全部需用語境語法的能力來處理。自然語言是略需語境語言 (Mildly Context-sensitive Languages)，因此也只要略需語境語法來處理。這是很重要的結果。因爲需用語境語法的剖析時間是指數時間 (exponential time)；即剖析所需時間和 2^n 成正比，而 n 則爲句子的長度。換句話說，每增一字剖析時間即增加一倍。這不但與人類處理自然語句的現象不符合，在電腦運算上也極不

18 某些否定句似乎可證明此類限制非純爲語法現象。

(i) 「是故鄉明月」不能對「露從今夜白」。

例句中顯示相應的結構只是一個語意的限制。雖然相對位置的字詞類不對稱，此句仍合句法。

經濟。免用語境語法的剖析時間僅和句子長度的三次方成正比，稱之爲多項式時間(polymial time)，比指數時間快多了，而且隨長度發生的變異不大。賈恤(Joshi 1985)認爲只要超出免用語境語法的範圍有限，剖析時間可以限制在多項式時間內。蓋志達與蒲倫(Gazdar and Pullum 1985)也指出絕大部份自然語言現象可以免用語境語法處理；其中的一大部份又可以更簡單，剖析時間更經濟的尋常語法處理。因此，在處理其他人工語言上所用的經濟有效的剖析運算方式大部分可用到自然語言上。簡而言之，雖然語言規範性質超出免用語境語言而包含於需用語境語言。需用語境語法所有不利於剖析運算的數學性質並不全然適用於自然語言。故以自然語言等同於規範上的需用語境語言所衍生的不利推論也無法成立。

本文對漢語的討論支持以上的看法。本文討論的幾個結構在其他漢語方言亦出現；可視爲漢語的共同特性。也就是說，漢語提供了自然語言中目前僅知的語法重覆實例。值得注意的是這些超出免用語境語言的結構均是數學上的複寫語言。所以好的漢語語法，除了要能處理複寫語言外，只要具免用語境語法的能力即可。根據賈恤(Joshi 1985)的定義及蓋志達與蒲倫(Gazdar and Pullum 1985)的討論，這樣的語法應該足以反映人類快速剖析了解自然語言的語句及小孩輕易習得母語語法這兩個事實。至於具複寫語言特性的成份是否是自然語言中唯一超免用語境語法的現象，則有待更進一步研究[19]。而寫出這麼一個語法，更是當前漢語語法學家面臨的挑戰。

致　　謝

本文曾在歷史語言研究所學術講論會中發表。感謝資訊所陳克健教授閱讀初稿並提出修訂意見。感謝二組同仁，特別是龔煌城教授，及其他與會同仁之建議。卡內基—美崙大學(Carnegie-Mellon University)的伯樂(Carl Pollard)教授在私下討論時提出了肯切的改進建議。文中若有謬失之處，文責由筆者自負。本文之成，曾獲工業技術研究院與中央研究院「中文語句剖析系統合作研究與開發計劃」及國科會「中文語句剖析的語法模式」(NSC-77-04088-E001-01)兩計劃之部份補助。特在此申謝。

19　施別(Shieber 1985)所討論的瑞士德語同時具複寫語言與 $a^m b^n c^m d^n$ 語言的特性。不過其分析是否需要比處理複寫語言更強的語法則不得而知。

參 考 文 獻

Bresnan, Joan, Ronald M. Kaplan, Stanley Peters, and Annie Zaenen

1982 Cross-serial Dependencies in Dutch. In *Linguistic Inquiry* Vol. 13. No. 4. pp. 613-623.

Chao, Yuen Ren

1968 *Language and Symbolic System.* Cambridge: Cambridge University Press.

1976a Notes on Chinese Grammar and Logic, In *Aspects of Chinese Sociolinguistics* pp. 237-249. Stanford: Stanford University Press.

1976b How Chinese Logic Operates. In *Aspects of Chinese Sociolinguistics* pp. 250-259. Stanford:Stanford University Press.

Chomsky, Noam

1963 Formal Properties of Grammars. In R. D. Luce, R. R. Bush, and E. Galanter eds. *Handbook of Mathematical Psychology,* Vol. Ⅱ. pp. 323-418. New York: Wiley.

Gazdar, Gerald

1981 On Syntactic Categories. In *The Psychological Mechanisms of Language* pp. 53-61. London: The Royal Society and the Birtish Academy.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan Sag

1985 *Generalized Phrase Structure Grammar.* Oxford: Blackwell and Cambridge: Havard University Press.

Gazdar, Gerald and Geoffrey K. Pullum

1985 Computaionally Relevant Properties of Natural Languages and Their Grammars. in *New Generation Computing.* Vol. 3. pp. 273-306.

Hoekesma, Jack

1987 Construction Types in Categorial Grammar. Paper Presented at

Penn Colloquium.

Hopcroft E. John and Jeffrey D. Ullman

1979 *Introductin to Automata Theory, Languages and Computation.* Reading: Addison Wesley.

Huang, C. T. James

1982 *Logical Relation in Chinese and the Theory of Grammar.* MIT Ph. D. Dissertation.

Joshi, Aravind

1985 Tree Adjoining Grammars. In David R. Dowty, Larui Karttunen and Arnold M. Zwicky eds. *Natural Language Parsing: Pycholinguistic, Computational, and Theoretical Perspectives.* pp. 206–250. Cambridge: Cambridge University Press.

Kac, Michael B.

1987 Surface Transitivity, *Respectively* Coordination, and Context-freeness. *Natural Language and Linguistic Theory,* Vol. 5. pp. 414-452.

Kac, Michael B, Alexis Manaster-Ramer and William C. Rounds

1987 Simultaneous-Distributive Coordination and Context-freeness. *Computational Linguistics.* Vol. 13. Nos. 1–2, pp. 25–30.

Kaplan, Ronald M. and Joan Bresnan

1982 Lexical-functional Grammar: A formal System for Grammatical Representation. In J. Bresnan ed. *The Mental Representation of Grammatical Relations.* pp. 173–281. Cambridge: MIT Press.

Peters, Stanley and Robert Ritchie

1973 On the Generative Power of Transformational Grammars. *Information and Control.* Vol. 18 pp. 483–501.

Pullum, Geoffrey K. and Gerald Gazdar

1982 Natural Languages and Context-free Languages. *Linguistics and Philosophy* Vol. 4. pp. 471–504.

Radzinski, Daniel

1987 Unbounded Syntactic Copying in Mandarin Chinese. Manuscript. Harvard University.

Shieber, Stuart M.

1985 Evidence Against the Context-freenes of Natural Language. *Linguistics and Philosophy* Vol. 8. pp. 333-343.

Tang, Ting-chi

1986 Syntactic and Pragmatic Constraints on the V-not-V Question in Mandarin. Presented at the 19th International Conference On Sino-Tibetan Language and Linguistics, Columbus. Also in Tang (1988) *Studies On Chinese Morphology and Syntax*. pp. 597-612. Taipei Student.

Wall, Robert

1972 *Introduction to Mathematical Linguistics*. Englewood Cliffs: Prentice-Hall.

黃　居　仁

On Formal Properties of Chinese

Abstract

Chu-Ren Huang

The study of mathematical properties of natural languages has both theoretical and applicational implications. In theoretical linguistics, a precise characterization of natural languages in terms of formal models, such as the Chomsky hierarchy, helps to capture the definition of possible natural languages and possible grammars for natural languages. Thus, the advantages and disadvantages of current grammatical theories can be compared and contrasted to motivate meaningful improvements. On the other hand, in natural language processing, knowledge of the formal properties of natural languages means the ability to chose and implement attested efficient parsing algorithms developed for corresponding formal languages.

As a first study of the formal properties of Chinese, this article determines the position of this language in the Chomsky hierarchy. Formal proofs are given to show that Chinese is neither a (type 3) regular set nor a (type 2) context-free language. Thus it is concluded that Chinese is supra-context-free. These formal proofs base on the closure properties of a type N language under substitution and under intersection with regular sets.

First, with central embedding relative clauses, it is shown that the intersection of Chinese with a well-defined regular set is not a regular set. Thus, according to the finite closure of regular sets under intersection, Chinese cannot be a regular set and it requires a grammar more complex than type 3 grammar.

Second, with three sets of data involving identical copying, it is

shown that the intersections of Chinese with well-defined regular sets are not context-free languages. The data discussed are A-not-A questions, interrogative sentential objects of 〔不管〕'to disregard,' and sentence-initial NP-not-NP 'regardless of NP.' Since context-free languages are closed under intersection with regular sets, Chinese cannot be a context-free language.

The proof of Chinese's being supra-context-free is significant as the first attested case of copying languages. Previously, the only generally accepted case of supra-context-free natural languages was Swiss German given in Shieber (1985), and the formal proof relies on homomorphism. The formal proof using identical copying of constituents of indefinite length in Chinese puts the fact that natural languages are supra-context-free beyond doubt. However, it is also true that the three sets of data requiring mechanisms stronger than context-free grammar all involve identical copying. Thus, Joshi's (1985) idea of 'mildly context-sensitive grammar' with limited supra-context-sensitive mechanisms and Gazdar and Pullum's (1985) suggestion of context-free based parsers for natural languages are still plausible.