

國語語言辨認中 詞群雙連語言模型的解碼方法

A Word-Class Bigram Approach to Linguistic Decoding in Mandarin Speech Recognition

林頌堅⁺ 簡立峰^{*} 陳克健^{*} 李琳山^{++*}

⁺國立台灣大學資訊工程學研究所

^{*}中央研究院資訊科學研究所

摘要

在國語語言辨認的語言解碼方法(linguistic decoding approach)中，字雙連語言模型(character bigram)和詞雙連語言模型(word bigram)是兩種最常被使用的方法。其中，詞雙連語言模型在描述語言現象上有較字雙連語言模型強的能力；然而若詞彙量較大時，它所需要估計的參數卻遠較字雙連語言模型多。因此若考慮在大詞彙、無限文句的國語語言辨認應用上，此二者皆有其適用上的限制。本文乃提出一個詞群雙連語言模型(word-class bigram)的語言解碼方法，這個方法是以對詞分群來大幅縮減參數量的大小，且又能接近詞雙連語言模型的辨認效果。分群方法是根據國語特殊的構詞特性來分群，比起其他西文常使用依據語法(syntactical)或語意訊息(semantic information)的分群方法不但計算簡單，且同樣具有分群的功用，而且詞群並不必須事先訓練或做詞類標記(part-of-speech tagging)即可決定，同時此法對新增詞及低頻率的詞也具某種程度的平滑(smoothing)能力。此外為達成即時計算的要求，我們針對構詞及格狀詞組搜尋提出二項技術來加快它們的速度。目前這個方法已實際應用在國語語言辨認上，實驗結果證實這個方法是相當實用且有效的。

第一節 緒論

在國語語音辨認的應用中，所謂語言解碼(linguistic decoding)的過程是指在一個由音節辨認(syllable recognition)後所獲得的格狀詞組(word lattice)或格狀字組(character lattice)中找出最可能的文句。傳統上，詞雙連語言模型(word bigram)或字雙連語言模型(character bigram)都是常用的方法[1,2,3,4]。基本而言，這兩種方法各有優缺點[2]：詞雙連語言模型在描述語言現象上有較強的能力；然而若詞彙量較大的話，它所需要估計的參數卻遠較字雙連語言模型多。若是考慮在大詞彙、無限文句的語音辨認應用上，這兩種方法都並不十分適合。因此，一個參數量比詞雙連語言模型小很多，但卻比字雙連語言模型有效的語言模型及快速解碼方法是非常需要的。

近年來盛行於西文的將詞分群以減少模型參數量的方法似乎是一個可行的方法[5,6]。然而此類方法大多是依據詞的語法或語意訊息來分群，由於需要做詞類標記(part-of-speech tagging) [5] 或是事先需要利用語料庫來抽取其中的語法或語意訊息做為分群上的依據[6]，所以分類的成本相當高，另外這類方法在面對新增詞大增時都可能必須重新分群訓練。因此本文根據中文的構詞特性提出一相當簡單的稱為「詞群雙連語言模型(word-class bigram)」的統計式語言模型。

這個模型的基本構想是中文具有相同起始字或結尾字的詞常有接近的語意，如一些複合詞或定量詞的中心語是在詞尾，像物理學、化學、科學，車、火車、機車，一個、十個、百個；而動補式動詞和動賓式動詞的中心語在詞頭，像打破、打壞、打爛、打棒球、打電話等。因此對所有的中文詞，我們可以考慮根據起始字和結尾字將它們分成若干群。分群的方法是對具有相同起始字的詞都歸成同一群，例如前述：打破、打壞、打爛等都歸成同一群 $S(\text{打})$ 。同樣的，對具有相同結尾字的詞，如前述：車、火車、機車等都歸成同一 $E(\text{車})$ 。因此，對於每一詞串(word sequence)，其機率估算方法，可以改以詞群的機率來計算；這種機率值的逼近方式相對於西文使用依據語法或語意訊息(syntactical or semantical information)的分群方法，不僅計算非常簡單，而且無須事先以語料庫訓練即可決定詞群，並且當新增詞大增時，詞群也不須重新認定。當然依

據這種方式，一些與語意差距甚大的詞也會歸於同一群，如：國中生、國語、國家等，因此在第二節裡我們提出更進一步的方法。根據我們對國語語音辨認的實驗發現，這種方法與字雙連語言模型相比較，因為它的詞群數目和中文字的數目一樣，所以它所需的參數量接近字雙連語言模型，但遠較詞雙連語言模型小。同時，此種方法在國語語音辨認上的效果比字雙連語言模型好。

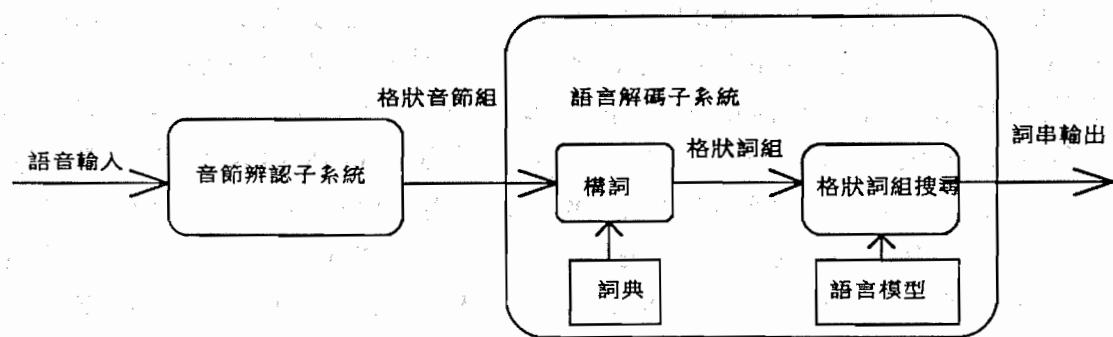
然而詞群雙連語言模型的語言解碼方法由於必須先構成格狀詞組，再從其中搜尋，構詞所花費的時間相當多。因此，我們特別建一個輔助表來及早刪除不可能成詞的音節串。如此一來，因為減少了多次不必要的詞典搜尋，使得構詞速度加快很多。另外我們從實驗觀察到，即使正確的詞不在最可能的一條詞串中被搜尋到，也很可能出現在前幾條詞串中。因此，若能使系統多搜尋幾條詞串輸出，如此有較多的詞可供選擇，一來便於人工修正；二來未來可加入語法或語意規則做為後處理，可以較詳細地過濾掉不合語法語意的文句。然而在多搜尋幾條詞串使正確的詞能出現的同時，為了不增加使用者和系統的負擔，我們也希望不正確的詞不會因此而增加太多。因此本文根據近來常用於連續語音辨認的樹狀一格狀最佳N條路徑搜尋法(tree-trellis N-best paths searching)[7]修改成適合我們應用的一個在搜尋時間和記憶體用量都是最佳化的格狀詞組的多條路徑搜尋法。根據我們實驗的結果，運用此方法搜尋五條詞串比搜尋一條詞串可以多對了10%的字，但只需多找40%的字，並且速度仍然相當快。

這個語言解碼方法已實際應用在國語語音辨認中，並獲得了很好的結果。另外，因為辨認的結果是以詞串的形式輸出而非僅以字串的形式輸出，此法對需要以詞串做處理單位的其他更高階的自然語言處理(natural language processing)，如語音瞭解(speech understanding)、語音機器翻譯(speech machine translation) 和資訊檢索(information retrieval) 等應用有相當大的助益。

第二節 詞群雙連語言模型

通常國語語言辨認系統可分為音節辨認子系統(syllable recognition subsystem) 和語言解碼子系統(linguistic decoding subsystem) 兩部分，其架構

如圖一所示[1,3,4]。我們所提出的詞群雙連語言模型就是一種語言解碼的方法，它包括了兩個過程：構詞(word formation) 和格狀詞組搜尋(word lattice searching)。所謂構詞是將一個由音節辨認所到格狀音節組(syllable lattice) 與詞典做匹配(matching) 而形成一個格狀詞組；而格狀詞組搜尋是根據語言模型(language model) 在所得到格狀詞組中搜尋一或多條可能的詞串做為輸出。



圖一 國語語音辨認系統架構圖

由於構詞和格狀詞組搜尋都要花費相當多的時間，尤其是當詞彙量非常大時，情形更是嚴重。因此在我們所提的語言解碼方法中特別加入一些輔助方法來加快它們的速度，這些方法在下一節中會有詳細的說明。此節中所要闡述的是前述我們所提語言模型—詞群雙連語言模型以及它的訓練方法。

機率計算

若產生一詞串， W_1, \dots, W_n 的機率為 $Pr(W_1, \dots, W_n)$ ，則

$$Pr(W_1, \dots, W_n) = Pr(W_1) \times Pr(W_2|W_1) \times \dots \times Pr(W_n|W_1, \dots, W_{n-1}) \dots (1)$$

由於上式所需的參數量過為龐大，前人的研究就作了一個馬可夫程序(Markov Process) 的假設，假設此詞串的產生是一個一階馬可夫模型(first-order Markov model) 的隨機程序(random process)，亦即每一個詞的產生只與它的前一個詞有關，因此：

$$Pr(W_1, \dots, W_n) \equiv Pr(W_1) \times \prod_{i=2}^n Pr(W_i | W_{i-1}) \quad \dots(2)$$

此處 $Pr(W_i | W_{i-1})$ 便是一個詞雙連語言模型的機率值。對每一個這樣的機率，如前節所說明的國語中複合詞的中心語在詞尾，因此大部分的詞，我們假設它可以用和前一詞的結尾字的條件機率(conditional probability)來逼近：

$$Pr(W_i | W_{i-1}) \equiv Pr(W_i | E(W_{i-1})) \quad \dots(3)$$

此處 $E(W_{i-1})$ 為具有與詞 W_{i-1} 相同的結尾字所有的詞所成的群。

若將每一個詞 W_i 視為由 $S(W_i)$ 和 $R(W_i)$ 兩部分組成，此處 $S(W_i)$ 為詞 W_i 的起始字部分；而 $R(W_i)$ 則是詞 W_i 除了起始字的部分，則由貝氏定理(Bayes' Theorem)，我們可以得到：

$$\begin{aligned} Pr(W_i | W_{i-1}) &\equiv Pr(W_i | E(W_{i-1})) \\ &= Pr(S(W_i)R(W_i) | E(W_{i-1})) \\ &= Pr(S(W_i) | E(W_{i-1})) \times Pr(R(W_i) | E(W_{i-1})S(W_i)) \quad \dots(4) \end{aligned}$$

再假設 $R(W_i)$ 的產生與前一詞 W_{i-1} 的結尾字 $E(W_{i-1})$ 無關，而又因為事件 $S(W_i)R(W_i)$ 產生的機率與事件 W_i 的機率相同，所以：

$$\begin{aligned} Pr(W_i | W_{i-1}) &\equiv Pr(S(W_i) | E(W_{i-1})) \times Pr(R(W_i) | S(W_i)) \\ &= Pr(S(W_i) | E(W_{i-1})) \times \frac{Pr(S(W_i)R(W_i))}{Pr(S(W_i))} \\ &= Pr(S(W_i) | E(W_{i-1})) \times Pr(W_i | S(W_i)) \quad \dots(5) \end{aligned}$$

式(5)便稱為詞群雙連語言模型的機率值。在此式中，前一項在具有相同結尾字的詞所成的集合出現之下，接連出現相同起始字的詞所成的集合的條件機率；而後一項為詞在具有與它相同起始字的詞所成的集合的條件機率。

參數量的評估

針對式(5)中 $Pr(S(W_i) | E(W_{i-1}))$ 的計算，我們需先對訓練語料斷詞(word segmentation)，接著統計事件 $E(W_{i-1})$ 和事件 $E(W_{i-1})S(W_i)$ 出現的頻率：

$$\begin{aligned} Pr(S(W_i) | E(W_{i-1})) &= \frac{Pr(E(W_{i-1})S(W_i))}{Pr(E(W_{i-1}))} \\ &\cong \frac{f(E(W_{i-1})S(W_i))}{f(E(W_{i-1}))} \quad \dots(6) \end{aligned}$$

同樣地， $Pr(W_i | S(W_i))$ 也是如此，但是由於事件 $W_i | S(W_i)$ 和事件 W_i 出現的機率相同，所以：

$$\begin{aligned} Pr(W_i | S(W_i)) &= \frac{Pr(S(W_i)W_i)}{Pr(S(W_i))} = \frac{Pr(W_i)}{Pr(S(W_i))} \\ &\cong \frac{f(W_i)}{f(S(W_i))} \quad \dots(7) \end{aligned}$$

由上面兩式我們來估計詞群雙連語言模型的參數量。每一個參數是由式(6)和式(7)兩部分組成。對式(6)而言，參數量的大小類似字雙連語言模型，但由於並不是每一個中文字都可做為詞的起始字或結尾字，因此式(6)的參數量應比字雙連語言模型稍少一些；而式(7)的參數量則與詞典大小有關，但相較於式(6)的參數量則顯得微不足道。因此，這個模型的參數量與字雙連語言模型參數量接近，但遠少於詞雙連語言模型的參數量。

其他問題的討論

以下我們針對語言模型常發生的問題來進一步討論。第一個問題是在詞雙連語言模型中由於要估計的參數量太大，所以有許多參數在語料庫中沒有出現或者是出現頻率太低，而使所估計的語言模型不準；針對這個問題由於詞群雙連語言模型所需估計的參數量比詞雙連語言模型小很多，而且分群具有平滑的效果，因此可以適度減輕這個問題。

其次這個語言模型由於是以詞的起始字和結尾字做為分群的依據，不需事先做詞類標記或從語料庫中抽取語法或語意訊息，因此不但運算簡單使訓練成本大大減低，同時若是有新增詞出現，由於可以自動分群並不需重新分群訓練。

此外這個語言模型具有比字雙連語言模型還強的語言現象描述能力，因此它的語言解碼的正確率比較高，這一點在第四節中將以實驗來佐證。但由於這個語言解碼方法必需先構詞，使得它花費的時間比較多，因此第三節我們將提出改進的方法。

第三節 構詞與格狀詞組搜尋

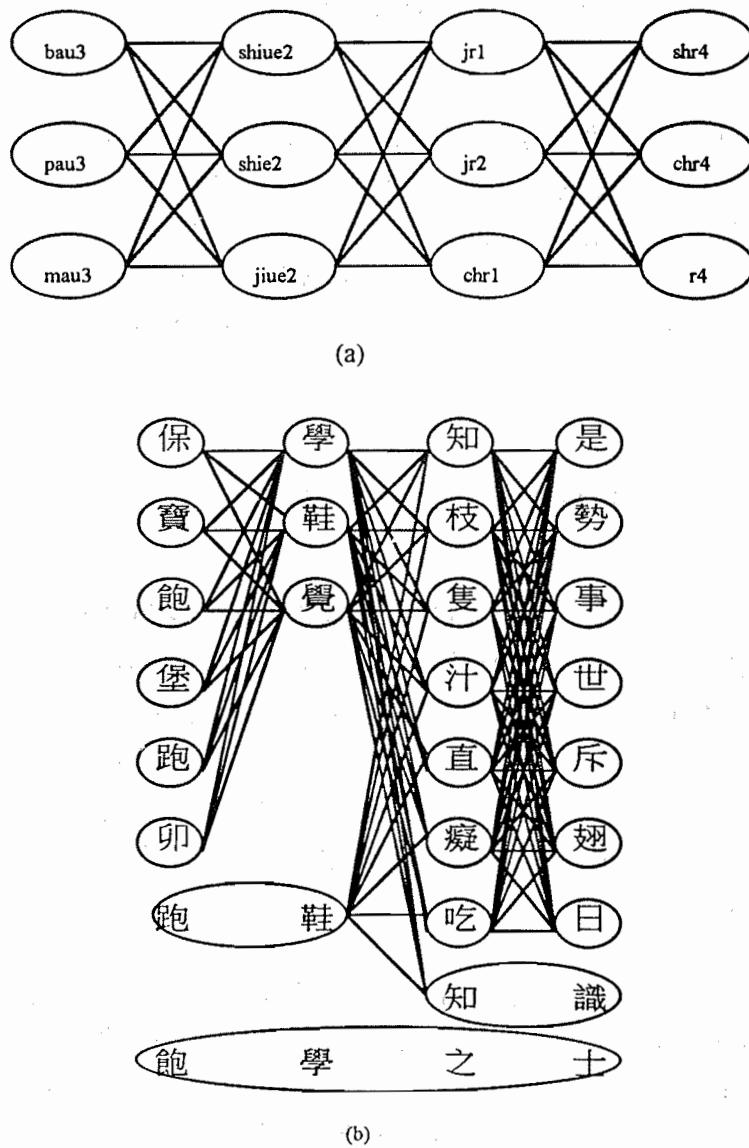
利用前節所提之語言解碼方法應用於國語語音辨認，當格狀音節組進入語言解碼子系統之後(圖一)，首先需經過構詞程序建構出一個相對應的格狀詞組，之後格狀詞組搜尋程序則在這個格狀詞組中搜尋出若干條可能的詞串輸出。由於這兩個程序都要花費相當的時間，所以針對這兩項速度的瓶頸，我們進而提出兩項技術使我們所提的語言解碼方法更為完整。

快速構詞法 (fast word formation)

構詞程序將某一給定的格狀音節組轉換成相對應的格狀詞組，如圖二所示。方法是查尋格狀音節組中每一可能的音節串(syllable sequence) 在詞典中有無對應的詞，每一音節串是由在格狀音節組有前後關係的音節所構成的鏈結(link) 串接起來。但由於在格狀音節組中會有很多不可能成詞的音節串（尤其是當音節串愈長時，愈會有這樣的情形發生），造成很多不必要的詞典查尋。若能降低這些不必要查尋的次數，將使構詞的速度加強不少。

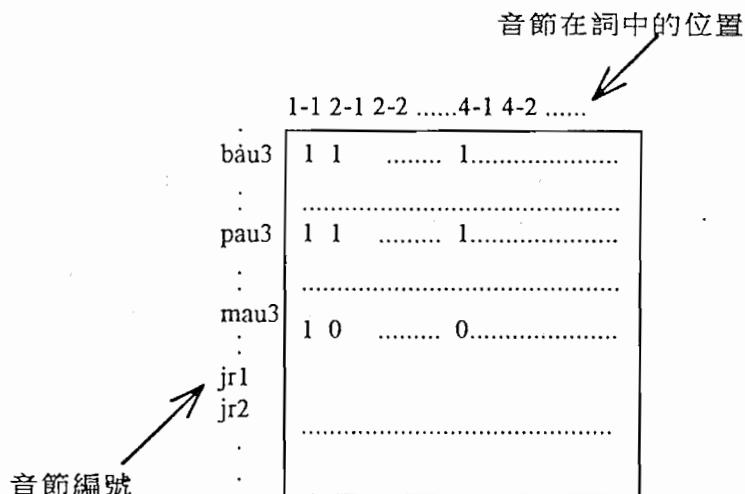
在這裡，我們的方法是使用一個輔助表來及早去除不可能成詞的鏈結：這個輔助表是由一組二元向量(binary vector) 組成；每一個二元向量代表每個音節在所有長度的詞中可能出現的位置：向量的第一個成分(component) 代表這個音節可否出現在單字詞，第二個成分代表這個音節可否出現在雙字詞的第一個字，第三個成分則是代表可否出現在雙字詞的第二個字，以此類推(圖三)。當要形成一個長度K的音節串 S_1, \dots, S_K 時，若在已形成長度(i-1)的鏈結 $S_1-S_2-\dots-S_{i-1}$ 串後，發現下一音節 S_i 在代表K字詞的第i個位置的二元向量的成分為0，即此音節 S_i 不能出現在K字詞的第i個字，則之後的鏈結便都不需形成。

舉例而言，如圖二(a)之格狀音節組，mau3¹並不會出現在4字詞的第一個字，所以像 mau3-shiue2-jr1-shr4, mau3-shiue2-jr1-chr4, mau3-shiue2-jr1-r4, ... 等鏈結便都不用形成。使用這個輔助表之後，大約減少了90%的詞典查尋，使得即時應用成為可能。



圖二 一個格狀音節組及其相對應的格狀詞組。(a)格狀音節組，(b)格狀詞組。

¹ 本文所採用之注音符號為教育部於民國七十五年頒布之國語注音符號第二式。



圖三 快速構詞法的輔助表示意圖

快速格狀詞組多路徑搜尋法(fast word lattice multiple path searching)

傳統使用維特比搜尋法(Viterbi's search)的國語語音辨認系統的語言解碼子系統，為了節省記憶體空間和搜尋時間，通常只搜尋一條最可能的文句，作為結束；而辨認錯誤之處則由使用者自行鍵入中文字來更正。但我們認為一個對使用者友善(user-friendly)的系統應該要減少使用者使用鍵盤更正的機會，所以需要一個能搜尋多條可能詞串的格狀詞組搜尋法。這樣一個能同時搜尋多條可能詞串的方法，除了可以提供給使用者更方便的環境外，未來這些詞串也才可以用需要較多運算的語法或語意規則來挑出最可能的詞串。同時在系統的建立時期，更可提供給設計者更多有關系統的能力和優缺點的資訊，進而加強系統的功能。

過去語言解碼系統只搜尋一條可能文句的原因主要是由於缺乏一個快速、使用記憶體少而有效的多條路徑演算法。在此，我們參考了莊、黃二氏為連續語音辨認(continuous speech recognition)所提出的樹狀一格狀最佳N條路徑搜尋法(tree-trellis N-best search) [7]，經過一些修改後，將其運用在格狀詞組的最佳N條可能的詞串的搜尋上。這個演算法是一個兩階段式的搜尋法：第一階段是正向的經過修改的維特比搜尋法(forward modified Viterbi's search)，第二階段是一個回向的A*演算法(backward A* algorithm)，而此回向A*演算法與其他A*演算法最大的不同就是其他A*演算法的優先值是一個預測值，而這一個回向A*演算法的優先值則是真

正經過此節點和回向未完成路徑的最佳分數，所以這個A*演算法是A*一可容許的(A*-admissible)[8]，因此它的記憶體用量和所花費的時間都是最經濟的。在我們的實驗中發現，若搜尋五條可能的詞串，可以比只搜尋一條詞串多對10%的字，而卻只要多找40%的字，這正是我們所要求的，能多辨認出一些正確的字，而多找出來的字又不造成太大的干擾。

第四節 實驗環境及初步實驗結果

為驗證所提出的語言解碼方法的效果，我們做了一系列在國語語音辨認應用的實驗，來比較它與傳統所採用的字雙連語言模型方法在辨認中文字上的效果。同時，我們觀察了一些辨認錯誤的結果，做為日後加強的參考。實驗結果顯示我們的系統在字的回叫率(recall rate)及精確率(precision rate)上都比傳統所使用的字雙連語言模型來得高。接下來在說明實驗結果之前，我們先介紹使用的實驗環境，包括音節辨認子系統、詞典以及訓練和測試語料。

音節辨認子系統

辨認結果	基底音節辨認	聲調辨認
前一個結果	93.87%	98.04%
前三個結果	99.51%	99.35%
前四個結果	100.00%	100.00%
前五個結果	100.00%	N.A.

表1 本實驗的音節辨認子系統的平均辨認率 N.A.代表NonAvailable

如前所述，語言解碼子系統從音節辨認子系統取得格狀音節組作為輸入，然後轉換出可能的文句輸出。實驗中，所使用的音節辨認演算法是根據金聲二號國語聽寫機的演算法[4]。此音節辨認子系統可分為兩部分：基底音節辨認(base syllable recognition) 及聲調辨認(tone recognition)。對每一個輸入音節的語音，基底音節辨認自408個基底音節中選5個最有可能的候選音節；而聲調辨認4個聲調(lexical tone)和1個輕聲調(neutral tone)中

選出3個候選聲調。然後，將這兩組結果加以組合，去除掉不可能的基底音節和聲調組合後，送至語言解碼子系統做為輸入。平均的音節辨認率如表1所示。

詞典及訓練和測試語料

本實驗中所使用的詞典是由中央研究院詞知識庫小組所研發[9]，但為配合語音辨認上的需要，我們做了一些修改，包括：1.由於太長的詞在日常生活中很少使用，所以我們將五字以上的詞刪除。2.目前專有名詞的處理是視做一串單字詞來處理；有些字並沒有出現在詞典的單字詞中，但它們有可能出現在專有名詞之中。為了這緣故，我們將所有的中文字都納入作為單字詞。經過這樣的修改之後，我們的詞典共包括八萬多個中文詞，其不同長度中文詞的個數如表2所示。

	詞數
單字詞	14052
雙字詞	48339
三字詞	11559
四字詞	10433
五字詞	583
總和	84966

表2 詞典中不同長度的中文詞數目

為了訓練我們所提出的語言模型，需要經過斷詞處理的大量中文文章做為訓練語料。此處我們的訓練語料亦為中央研究院詞知識庫小組所發展[10]；共包括三家主要的報紙，即中國時報、聯合報及自由時報，約400萬個中文字。相信這對字雙連語言模型以及詞群雙連語言模型的訓練大致是足夠的。

因為所需語料量相當大，所以若以人工來斷詞的話，不但是件繁瑣的工作，而且要花費相當多的時間。因此，勢必要採用自動斷詞系統。在我們的實驗裡，採用的是陳、劉二氏所提出並發展的斷詞系統[11]。這個斷詞系統基本上是一個字串匹配演算法(string matching algorithm)加上

六條解決當某一段字串可以被斷成一種以上的詞串的情況的經驗法則。它的成功率可達99.77%，也就是說在訓練語料中仍有少許的雜訊(noise)。

而實驗所用的測試文章則是從三種不同文體的題材中隨機挑選出來，共有四篇從報紙中選取的新聞(共2309字)、一篇選自天下雜誌的短文(1548字)，以及一篇短篇小說「明還」(970字)(選自「人子」，鹿橋著，台北市，遠景出版公司出版)，這些都是由不包括訓練語料的語料庫中隨機抽取出來的。這三種文體相信已包含了日常生活中常用的詞彙和句型，應該可以瞭解我們所提出的方法的效果。

回叫率及精確率的定義

在本實驗中，字辨認的效果是由回叫率(recall rate)及精確率(precision rate)來評估，其定義如下：

假設 C_j 為測試文句的第 j 個字

H_{ij} 為經搜尋後第 i 條可能詞串的第 j 個字

則在搜尋 i 條可能詞串後，回叫率 R_i 的定義為：

$$R_i = \frac{\sum_{\text{for all sentences}} \sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \cap \{C_j\} \right|}{\sum_{\text{for all sentences}} n_s} \times 100\%$$

而精確率 P_i 的定義為：

$$P_i = \frac{\sum_{\text{for all sentences}} \sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \cap \{C_j\} \right|}{\sum_{\text{for all sentences}} \sum_{j=1}^{n_s} \left| \bigcup_{k=1}^i \{H_{kj}\} \right|} \times 100\%$$

此處， n_s 為測試文句的長度，而 $|A|$ 為集合 A 的元素個數。

上面兩式中，分子部分是對文章中每一句的每一字從搜尋到的第一條詞串到第*i*條詞串所辨認的字做聯集，然後與正確的字交集。因此，若正確的字不出現在辨認出來的字的聯集中，則為空集合，亦即集合元素個數為0，反之元素個數為1，然後再總和整篇文章。 R_i 的分母部份是文章的長度； P_i 的分母部份是所有搜尋到的字數。所以回收率是辨認正確的字數佔整篇文章的比率；而精確率是正確的字數佔全體找出字數的比率。由於是斷開音節(isolated syllable)的國語語音辨認，所以不會有插入(insertion)和刪除(deletion)字的情況發生，因此第一條搜尋出的可能詞串的字數與所輸入文句的字數相等，也就是在搜尋第一條詞串後，其回叫率 R_1 應該等於精確率 P_1 。

舉例而言，如果要辨認的句子是「維持現有名額」，而經搜尋後得到第一條和第二條詞串分別是「維持現有名額」和「維持現有名額」，則此時回叫率和精確率的計算如下：在此 $n_s=6$ ，而第一條詞串辨錯兩個

字，所以 $\sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \cap \{C_j\} \right| = 4$ ， $i=1$ ，而 $\sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \right| = 6$ ， $i=1$ ，因此回叫率 $R_1 = 4/6 * 100\% = 66.67\%$ ，精確率 $P_1 = R_1 = 66.67\%$ ；而第二條詞串多辨出兩個字，所以 $\sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \cap \{C_j\} \right| = 6$ ， $i=2$ ，而 $\sum_{j=1}^{n_s} \left| \left(\bigcup_{k=1}^i \{H_{kj}\} \right) \right| = 8$ ， $i=2$ ，因此回叫率 $R_2 = 6/6 * 100\% = 100\%$ ，精確率 $P_2 = 6/8 * 100\% = 75\%$ 。

在傳統語音辨認實驗之中，通常正確率是指第一條可能詞串的回叫率。但正如第三節所指出的我們認為一個對使用者友善(user-friendly)的系統，應該讓使用者減少鍵入的次數，除了能搜尋多條可能的詞串之外，我們也希望系統所找出的正確字數愈多愈好而不需要的字不會因此增加太多。換言之，除了回叫率之外，精確率也是一項重要的評估標準。底下的實驗結果顯示，我們所提出的方法確實符合我們的要求。

測試結果

正如前面所提到的，對每一段音節的語音輸入，音節辨認子系統找出五個候選音節及最多三個候選聲調作為語言解碼子系統的輸入。此外，對每一輸入文句的語音，語言解碼子系統針對不同的應用可以找出不同數目的可能文句輸出。因此，可以形成很多組不同的實驗參數。底

下我們說明其中一些值得探討的現象，其結果顯示在表3至表6，表中包括了兩種不同的音節辨認結果及兩種語言解碼方法辨認三種不同文體的測試文章所得到的回叫率及精確率。

模型	新聞		雜誌		短篇小說	
	回叫率	精確率	回叫率	精確率	回叫率	精確率
字雙連語言模型	83.61%	83.61%	83.20%	83.20%	77.73%	77.73%
詞群雙連語言模型	87.93%	87.93%	87.92%	87.92%	77.73%	77.73%

表3 兩種不同的語言解碼方法測試三種不同文體的平均回叫率和精確率，實驗參數為取一個正確的聲調和五個候選基底音節及只找一條可能的詞串。

模型	新聞		雜誌		短篇小說	
	回叫率	精確率	回叫率	精確率	回叫率	精確率
字雙連語言模型	90.08%	64.38%	92.25%	64.50%	86.39%	59.18%
詞群雙連語言模型	93.40%	66.85%	95.16%	67.72%	86.29%	60.65%

表4 兩種不同的語言解碼方法測試三種不同文體的平均回叫率和精確率，實驗參數為取一個正確的聲調和五個候選基底音節及找五條可能的詞串。

模型	新聞		雜誌		短篇小說	
	回叫率	精確率	回叫率	精確率	回叫率	精確率
字雙連語言模型	82.75%	82.75%	83.07%	83.07%	77.42%	77.42%
詞群雙連語言模型	87.11%	87.11%	87.73%	87.73%	77.32%	77.32%

表5 兩種不同的語言解碼方法測試三種不同文體的平均回叫率和精確率，實驗參數取三個候選聲調和五個候選基底音節及只找一條可能的詞串。

模型	新聞		雜誌		短篇小說	
	回叫率	精確率	回叫率	精確率	回叫率	精確率
字雙連語言模型	89.31%	63.82%	92.12%	64.41%	86.19%	59.08%
詞群雙連語言模型	92.67%	66.18%	95.03%	67.57%	87.01%	61.20%

表6 兩種不同的語言解碼方法測試三種不同文體的平均回叫率和精確率，實驗參數為取三個候選聲調和五個候選基底音節及找出五條可能的詞串。

從這些表中，我們可以發現所提出的詞群雙連語言模型語言解碼方法所得到的辨認結果較傳統用字雙連語言模型的方法還要好。同時，可以發現在三種不同文體之中，以「短篇小說」的辨認結果最差，相信這是由於我們用新聞作為訓練語料，而這篇短篇小說的文體和詞彙與新聞相去甚遠的緣故，若是訓練語料中能包括這種文體，結果應該不會太差。

錯誤結果觀察

在這裡，我們觀察了一些詞群雙連語言模型語言解碼辨認錯的結果，這些錯誤可分為五類。第一類是由於雙連語言模型的影響。這類錯誤的主要原因是辨認的前後詞在訓練語料中接連出現的次數較少，使得這兩個詞一起出現的機率很低。

第二類錯誤是詞在詞群的相對頻率太低而引起。如jia1 ru4會被辨認成「掐住」的原因，是因為「加入」在S(加)中的相對頻率太低。

第三類錯誤是長詞遮蓋短詞，例如：he2 li3被辨認成「合理」而不是「河 裡」，而shr2 jian1被辨認成「時間」而不是「十 間」。因為同樣字數的文句中，長詞比短詞少乘了好幾項機率，所以長詞的機率比較容易來得高。

第四類錯誤則是專有名詞和簡稱。由於每天都會有許多專有名詞產生及消失，所以我們不可能也沒有必要將所有的專有名詞收入詞典中。但由於我們的測試資料包括了新聞類的文體，所以錯誤的專有名詞和錯誤的簡稱情形相當嚴重。

最後一類錯誤是同音且語意接近的詞，例如：「他」、「她」、「它」。要校正這種錯誤，只靠語法的統計訊息是不夠的，還要加入更高層的語言訊息，如語意(semantic)和語用(pragmatic)等。有些錯誤為同音而且同義的詞，如「充分」、「充份」和「保母」、「保姆」、「褓姆」等。這種錯誤嚴格來說並不能算是錯誤，因為事實上或許根本沒有標準答案。

第五節 結論

在國語語言辨認的語言解碼方法中，字雙連語言模型和詞雙連語言模型是兩種最常被使用的方法。其中，詞雙連語言模型因較描述語言現象的能力，所以辨認效果較好，然而所使用的參數量較大；相反的，字雙連語言模型的參數量較小，而辨認效果卻較差。在大詞彙、無限文句的國語語音辨認的應用上，這兩個方法都有其適用上的困難。本論文提出一個詞群雙連語言模型的語言解碼方法。這個方法包括一個新的語言模型—詞群雙連語言模型並且提出兩項技術來克服速度上的瓶頸：以一個輔助表來及早去除不可能的成詞的音節串，使構詞的速度加快；而以一個兩階段式的格狀詞組多路徑搜尋法用最經濟的記憶量和時間找到最佳的N條可能詞串輸出。

我們所提出的語言模型是一種以對詞分群的方法，用來大幅減少模型的參數量，且仍然接近詞雙連語言模型的效果。這種分群的方法是根據國語特殊的構詞特性，用詞的起始字和結尾字做為分群的依據，這種根據中文特有語彙訊息的方法不僅具有其他分群方法在節省記憶體空間及對低頻的詞平滑化的優點，同時計算上相當方便；經實驗證明辨認結果也很理想。

實驗結果顯示，在八萬多個詞彙和400萬個字的訓練語料的系統中，詞群雙連語言模型對三種測試文體的辨認結果都比字雙連語言模型好。因此，我們所提出的方法是相當實用而有效的。此外，若將這個語言解碼子系統的輸入和構詞程序加以修改，這個語言解碼方法便可適用於其他類似的應用，如注音輸入法的音轉字方法及光學中文字形識別的後處理。

致謝

在此感謝台大資訊所楊燕珠同學、張元貞同學和楊榮荌同學幫忙撰寫所需程式及提供給我們寶貴的意見，以及中央研究院詞知識庫小組提供語料庫和詞典。

參考文獻

- [1] L. S. Lee, C. Y. Tseng, H. Y. Gu, K. J. Chen, F. H. Liu, C. H. Chang, S. H. Hsieh, C. H. Chen, "A Real-time Mandarin Dictation Machine for Chinese Language with Unlimited Texts and Very Large Vocabulary," ICASSP'90, pp.65-68.
- [2] H. Y. Gu, L.Y.Tseng, and L. S. Lee, "Markov Modelling of Mandarin Chinese for Decoding the Phonetic Sequence into Chinese Characters," Computer Speech and Language, Vol. 5, NO. 4, Oct. 1991, pp.363-377.
- [3] L. S. Lee, C. Y. Tseng, H. Y. Gu, F. H. Liu, C. H. Chang, Y. H. Lin, Y. Lee, S. L. Tu, S. H. Hsieh, C. H. Chen, "Golden Mandarin (I) - A Real-time Mandarin Dictation Machine for Chinese Language with Very Large Vocabulary," to appear in IEEE Trans. on Speech and Audio Processing, Vol.1, NO. 2, Apr 1993.
- [4] L. S. Lee, et. al., "Golden Mandarin (II) - An Improved Single-chip Real-time Mandarin Dictation Machine For Chinese Language with Very Large Vocabulary," ICASS'93, pp. 503-506.
- [5] A. Derouault and B. Merialdo, "Natural Language Modeling for Phoneme-to-Text Transcription," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, NO. 6, Nov. 1986, pp.742-749.
- [6] P. F. Brown, V. J. Della Pietra, D. V. deSouza, J. C. Lai, and R. L. Mercer, "Class-Based n-gram Models of Natural Language," Computational Linguistics, Vol. 18, NO. 4, 1992, pp.467-479.

- [7] F. K. Soong and E. Huang, "A Tree-Trellis Based Fast Search for Finding the N-Best Sentence Hypotheses in Continuous Speech Recognition," ICASSP'91, S10.5, 1991, pp.705-708.
- [8] N. J. Nilsson, Principles of Artificial Intelligence. Tioga Publishing Company, California, 1980.
- [9] 黃瑞珠, 詞彙檔案編輯及維護系統使用手冊. 中央研究院詞知識庫小組, 技術手冊編號CKIP-93-07, 1993.
- [10] C. R. Huang and K. J. Chen, "A Chinese Corpus for Linguistic Research," Proceeding COLING'92, Vol. 4, 1992, pp. 1214-1217.
- [11] K. Chen and S. Liu, "Word Identification for Mandarin Chinese Sentences," Proceedings of COLING'92, Vol. 1, 1992, pp.23-28.