國立交通大學

電信工程學系碩士班

碩士論文

台語語音辨識與文字處理之研究

Studies on Taiwanese Speech Recognition and Text Analysis

研 究 生：王文德

指導教授：陳信宏　博士

中華民國九十三年七月

台語語音辨識與文字處理之研究

Studies on Taiwanese Speech Recognition and Text Analysis

研 究 生：王文德　　Student： Wen-De Wang

指導教授：陳信宏 博士　　Advisor： Dr. Sin-Horng Chen

A Thesis
Submitted to Department of Communication Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the Degree of
Master of Science
In
Electrical Engineering
June, 2004
Hsinchu, Taiwan, Republic of China

中華民國九十三年七月

台語語音辨識與文字處理之研究

研究生：王文德　　指導教授：陳信宏　博士

國立交通大學電信工程學系碩士班

中文摘要

本論文探討台語語音處理的兩個問題：語音辨識與文句分析，在台語語音辨識方面，我們建立聲韻母次音節隱藏式馬可夫模型，作連續語音音節辨識，在多語者的情況下，男女生之音節辨識率分別為42.67%及 47.33%；在台語文句分析方面，我們使用詞典依長詞優先之原則進行斷詞，對於台語文字表示法不統一而引起的斷詞混淆問題，我們藉建立音節與字的對應表，將詞典之詞展開，來改善斷詞之正確率。

關鍵詞：台語語音辨識、隱藏式馬可夫模型、台語文句分析、斷詞

Studies on Taiwanese Speech Recognition and Text Analysis

Student: Wen-De Wang Adviser : Dr. Sin-Horng Chen

Department of Communication Engineering

National Chiao Tung University

ABSTRACT

In this thesis, two tasks of Taiwanese language processing are studied. First, the task of Taiwanese speech recognition is exploited. A set of *initial* and *final* sub-syllable hidden Markov models (HMMs) is constructed for continuous base-syllable recognition. Syllable accuracy rates of 47.33% and 42.67% were obtained for multi-speaker female and male speech recognition, respectively. Then, the task of Taiwanese text analysis is studied. The problem of word tagging ambiguity due to the non-standardization of Taiwanese written form is exploited. A method to expand the representations of words in the lexicon by using a syllable-to-character mapping table is proposed. Experimental results confirmed the effectiveness of the proposed method on improving the performance of word tagging.

Keywords: Taiwanese speech recognition, hidden Markov model, Taiwanese text analysis, word tagging

目錄

表目錄

圖目錄

第一章 緒論

電腦的出現，使人類的生活變的多彩多姿也更加的便利，也將人類推向了一個充滿幻想的世界。而人類與電腦之間的溝通除了傳統的鍵盤與稍具人性化的滑鼠外，還有另外的選擇，那就是利用人類最自然也是最原始的工具--聲音，畢竟聲音是人與人溝通最直接的橋梁，如果能將聲音運用在人和電腦的溝通上，相信對於人類的生活一定會增加更多的樂趣與便利。在以前的時代或許這只是個幻想，不過，在語音辨識和語音合成的技術出現後，這就不再是空談了，而是一件可以實現的事。其中語音辨識的功能主要是讓電腦能聽懂人類所說的話，相對地，人也得要聽的懂電腦所說的話，而這就是語音合成的功能。

1.1.研究動機

台語(閩南語)是台灣人最常使用的方言，相信所有台灣人都會說一兩句的台語，在文字方面，雖然台語並無一套真正屬於自己的文字，但卻可藉由漢字與拼音的組合來表達台語真正的意思，例如台語的”sian3”和”sam3”在國語方面卻只能用”打”來替代，但是兩者意義卻不同，”sian3”是屬於力道較小的打而”sam3”是屬於力道較大的打【1】，因此，我們希望電腦要能聽的懂台語的語言和看的懂台語的文章，如此才能與人類溝通的更順暢。由於本實驗室之前並沒有人做過台語語音辨識相關的研究，因此在本論文裡我們將對台語語音辨識方面做較初步的研究、以及嘗試解決台語文章寫法混亂的問題，例如：”台灣儂”和”台灣人”皆為國語”台灣人”的意思。

1.2.研究方向

本論文包含了台語語音辨識以及台語文字處理兩個問題，在語音辨識方面，我們建立屬於台語本身的聲學模型(acoustic model)並利用此聲學模型來進行單音節的辨識，而在文字處理方面，將會改進我們現有的詞庫，與斷詞的方法，並描述在處理台語文字時所會遭遇到的問題，以及利用現有的資源企圖降低因台語文字混亂對斷詞器造成的斷詞不正確的影響。

1.3.章節概要

第一章 緒論：介紹研究動機，方向以及章節概要。

第二章 台語語音辨識：建立台語聲學模型，分析辨識後的結果。

第三章 台語文字分析：描述在寫台語文章時選字上的問題，以及如何利用詞典來降低因寫法混亂所造成的問題。

第四章 結論與未來展望：說明本論文的結論及未來改進方向。

第二章 台語語音辨識

2.1.簡介

目前國內台語語音辨認系統已有長庚大學呂仁園教授的不特定語者大辭彙華台雙語辨認系統，在一萬詞時其辨識率為 73.50%而七萬詞時其辨識率為 58%【2】。在本論文中我們將建立台語的聲母(initial)和韻母(final)的聲學模型(acoustic models)，並利用此模型做台語基本單音節之語音辨識。

2.2.台語的語音特性

台語和國語一樣都是聲調語言，音節由聲母、韻母和聲調(tone)所組成，不同的是台語的基本音節為 877 個，較國語多，其列表詳見附錄一；另外，台語的聲調共為八種，其中二、六聲已合併而只剩七種，也較國語的聲調多，各聲調之特徵及例字如表 2-1 所示，其典型基頻軌跡(pitch contour)如圖 2-1 所示。

表 2-1 台語八聲例表

聲調	台文字	羅馬拼音
一聲(陰平)	衫	saN
二聲(陰上)	褚	te2
三聲(陰去)	褲	kho3
四聲(陰入)	闊	khoah
五聲(陽平)	人	lang5
六聲(陽上)	矮	e2
七聲(陽去)	鼻	phiN7
八聲(陽入)	直	tit8

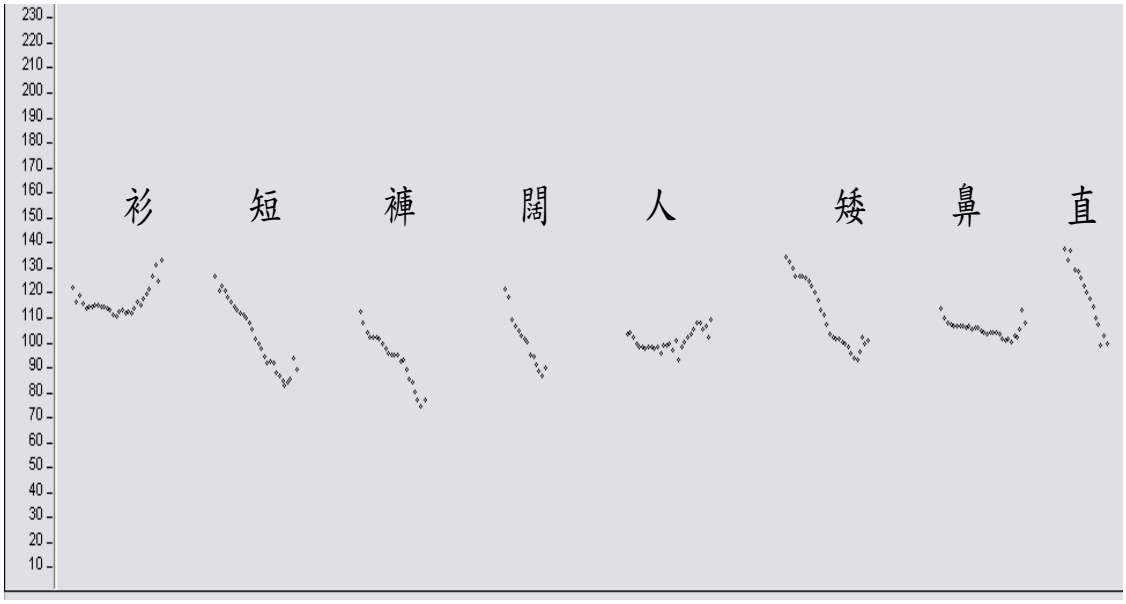

圖 2-1 台語八聲調之基頻軌跡

台語之聲母包含空聲母共 18 類，依其發音特點可分五類：(1) 脣音：p,ph,b,m，(2) 舌尖非齒音：t,th,l,n，(3) 舌根音：k,kh,g,ng，(4)舌尖齒音：ch,chh,s,j，(5)喉音：0(空聲母),h,。

台語之韻母包含空韻母共 84 類，其中又可分為基本韻母、入聲韻母、鼻化韻母與入聲鼻化韻母。基本韻母泛指不屬於其他三類之韻母，入聲韻母其符號特性是以 p,t,k,h 結尾，其中-h 韻尾在連讀時會失去，在個讀時則保留，因此個讀的發音是本音；鼻化韻母其符號特性是以 N 結尾，而入聲鼻化韻母其符號特性為 hN 結尾之韻母。

在台語中聲調扮演著相當重要的辨詞功能，而輕聲調在台語裡是較特殊的聲調，底下我們將介紹輕聲的特性【3】。輕聲的發音與第三聲很類似，主要是靠前後音節間的長短來分辨。而輕聲主要在多語詞之間才有辨語的功能，因此不將之歸於第三聲。輕聲化規律條件有二：(1) 能輕聲化的音節只出現在主要語法範疇 VP、NP、S 等的末尾；(2) 關連到數量語或代名詞時，只有在無語意重點時才可輕聲，因此產生了以下的功能：

1. 標誌出主要句法單位 NP，VP，S 的界線

   例如：

   a.走= 出來　　VP　　跑出來(國語意)

   b.走出來外口　　VP　　跑到外頭來(國語意)

   其中 a 句的出來為輕聲，因是在 VP 末尾。

2. 標誌語意重點

   例如：

   a.讀= 兩本　(重點在動詞，強調讀書的動作)

   b.讀兩本　　(重點在數量語，強調已讀了兩本書)

3. 分辨語詞

   例如：

   a.後= 日　　後天(國語意)

   b.後日　　改天(國語意)

4. 標誌虛詞所在

   台語雖不是所有的虛詞都輕聲，但是所有輕聲出現的情形都是 NP，VP，S 末尾的虛詞。這些虛詞可能是詞尾、詞組尾、或是句尾。

由以上的例子可發現，不論是否該讀輕聲其寫法皆相同，接下來我們將介紹目前大多數對於該唸輕聲詞的寫法：(1) 以等號或 “--” 標示出之後的詞應讀輕聲，例如：後=日；後--日；(2) 以 “0” 表示為輕聲調，例如：chhut-lai0 (出來)。

2.3.台語音節辨識系統基本架構

在進行語音辨識之前，我們先介紹辨識系統的基本架構，它的方塊圖見圖 2-2，主要包含三個部份：(1) 特徵參數擷取；(2) 聲學模型訓練；(3) 辨識比對。

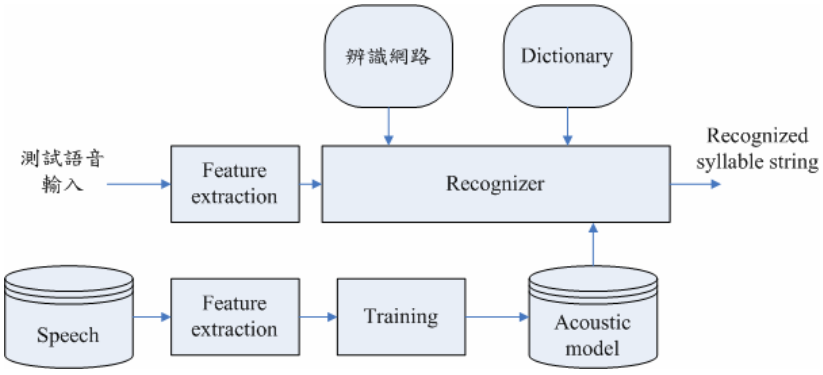

圖 2-2 語音辨識基本架構

各個方塊之功能說明如下：

- **Feature extraction**：對音訊做處理，求出能表示此語音的特徵訊息，以作為語音訓練及辨識之參數。
- **Training**：利用特徵參數與隱藏式馬可夫模型(Hidden Markov Model，HMM)求得一組聲學模型。
- **Dictionary**：在論文中辭典中僅是基本音節，所以此檔案記錄了聲學模型與基本音節的對應關係。在辨識時可經由查尋此檔案的動作來將辨識後的聲學模型符號轉成基本音節的拼音符號。
- **辨識網路**：作為辨識時所依據的搜尋網路，目前的辨識網路並無加任何限制。

2.4.資料庫的介紹

在進行語音辨認之前，首先我們必須收集足夠的語料來進行辨認模型的訓練，而這些語料都必須要有其相對應的標音，但並不是人人都會正確的說台語或是標示台語的拼音，所以對於語料的收集及標示並不容易。底下將介紹我們所擁有的語料庫以及對這些語料庫的處理。

1. Database1:含男女兩人以自然流利朗讀之大型 database，所錄製之內容是為故事類，錄製之方式是將文章分成許多小段分別錄製而成，最長段落其字數約為 250 個字，原來作為訓練 TTS 之用，其錄音之 sampling rate 為 20 kHz，我們將其 down sampling 為 16 kHz，以配合其他語料。
2. Database2: 由 99 人所錄製之語料，其 sampling rate 為 16 kHz，音檔共有 1034 個，文字檔共有 1033 個，未處理前文字共有 108169 字，在 99 人的語料中有些人讀相同的文章，相異文章的數目只有 57 篇。此資料庫有著比較多的問題說明如下：文字檔與音檔之對應關係不完全正確、語者間音量變化極大且有 amplitude saturation 的問題 (見圖 2-3)、以及含非台語發音。對於以上的問題我們將做以下的處理：(1) 以人工的方式修正其音檔與文字檔之對

應關係；(2) 先利用程式將 amplitude saturation 之音檔找出後，再以人工的方式聽此音檔，並留下此音檔可用之部份，也將對應之文字檔修改之；(3) 先利用程式列出拼音有非 877 音之文字檔，並配合音檔與辭典以人工的方式修改之，而非台語發音部分，將其作標記不使用。

3. Database3:由男十人及女八人錄音，是為補足前述語料以能充分訓練聲學模型，以及作為測試語料之用，內容皆為句子及短詞，取樣頻率為 16 kHz，文字取自“臺語通用會話：一套讓你能說出台語美辭雅語的教材.第二集【4】”及”一分鐘台語單字速成【5】”。

語料整理之後的結果列於表 2-2。

表 2-2：各語料之統計資料

	Database1		Database2		Database3	
音檔格式	無檔頭音檔(pcm)，取樣率為 16 kHz					
性別(人數)	男(1)	女(1)	男(46)	女(53)	男(10)	女(8)
總字數	23633	22010	30488	42720	4460	4460

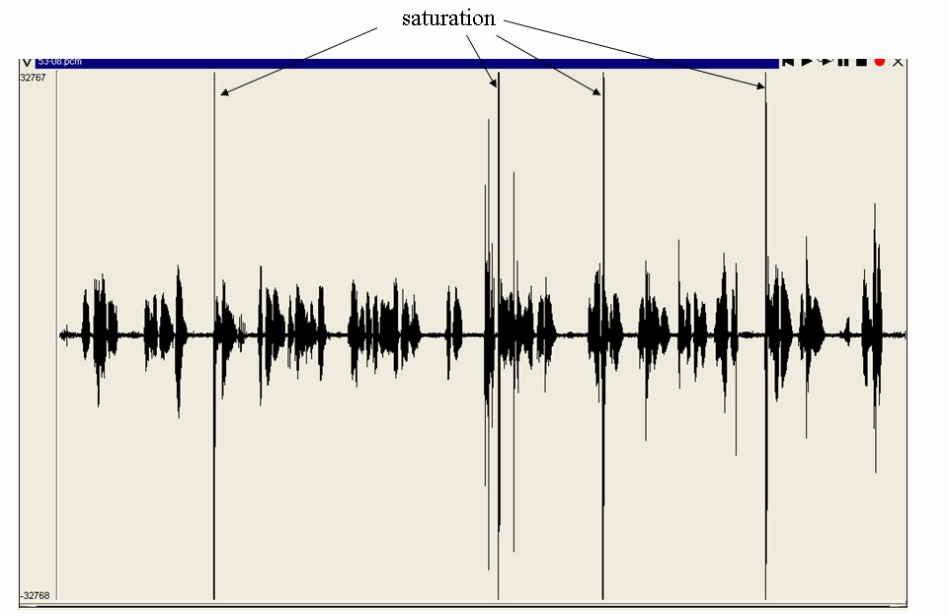

圖 2-3 amplitude saturation 的現象

2.5.特徵參數擷取

將原始的語音訊號經數位化後，雖然可直接拿來作辨識之用，不過由於資料量過於龐大，因而造成處理速度的緩慢，而特徵參數擷取是對音訊做處理，求出能表示此語音的特徵訊息，再利用所擷取出的特徵向量序列(feature vector sequence)，配合 HMM，來練訓聲學模型，所訓練出來的聲學模型即為往後要來做語音辨識的參考模型。在語音辨識之前要先設計出辨識網路，在辨識時除了需要之前所訓練出的聲學模型外，還必須配合維特比搜尋(Viterbi search)或光束搜尋(Beam search)，在辨識網路中進行比對搜尋，找出最佳之路徑，並由此對應出聲學模型，並利用其組合產生說話者所發音的拼音符號序列。在此我們採用語音界廣泛使用的梅爾刻度之倒頻譜係數(Mel-Frequency Cepstral Coefficients，MFCC)，經統計錄音者的說話速度為 3~4 字/秒因而將 Delta window size 與 Delta delta window size 設為 3，在本論文裡我們將取 38 維 MFCC 參數當特徵參數，相關之設定見表 2-3。

表 2-3 特徵參數設定

取樣頻率	16 kHz
預強濾波器	$1-0.97Z^{-1}$
視窗形式	Hamming window
音框長度(Frame size)	32ms
音框平移(Frame shift)	10ms
Filter bank	22
Feature vector	MFCC_E_D_A_N_Z
Delta window size	7
Delta delta window size	7

其中符號 MFCC_E_D_A_N_Z 之意義為 12 維 MFCC，13 維 Delta MFCC 和 13 維 Delta delta MFCC，並且做 Cepstral Mean Normalization。

2.6.聲學模型

將台語基本音節分解成聲母和韻母，並依據聲母右相關韻母的方式將 initial 分成 119 類見附錄一，並利用 HMM 來建立台語的聲學模型，因此我們所需要訓練的模型個數共為 205 類。initials 和 silence model 我們使用三個 states 的 HMM models 來建立，finals 則用五個 states 的 HMM models 來建立。short pause (sp) 是 syllable 與 syllable 之間的 silence，其 model 是一個 state 的 HMM model，採用與 silence model 中間的 state 共用。

在訓練的過程中，由於本實驗室之前並沒有人做過台語語音辨識之相關研究，因此並無較好的 *initial model* 也因此無法自動地獲得音檔之切割位置，在本論文裡我們採用 HTK【6】裡所提及的方式 *flat start*，其方法是認為開始時所有的 HMM model 參數皆令為一樣，再利用 Baum-Welch 的方式更新各個 HMM model 的參數，雖然此法會花費很多的時間，但我們所擁有的語料並不多因此不用花費太多的時間即可得到不錯的模型，其 *initial model* 建立過程如下：

以類似 uniform segmentation 的方式對訓練語進行切割並訓練出一個 global HMM model 其 mean 和 covariance 等於 global speech mean 和 covariance，並指定每一個 HMM model 皆為此 global HMM model。

2.7.辨識網路

由於我們無大量的台語文字庫，故無法獲得台語的語言模型 (Language model)，因此目前我們的辨識網路並沒有加任何的限制，即為任何音節皆可接任何音節，圖 2-4 為示意圖。

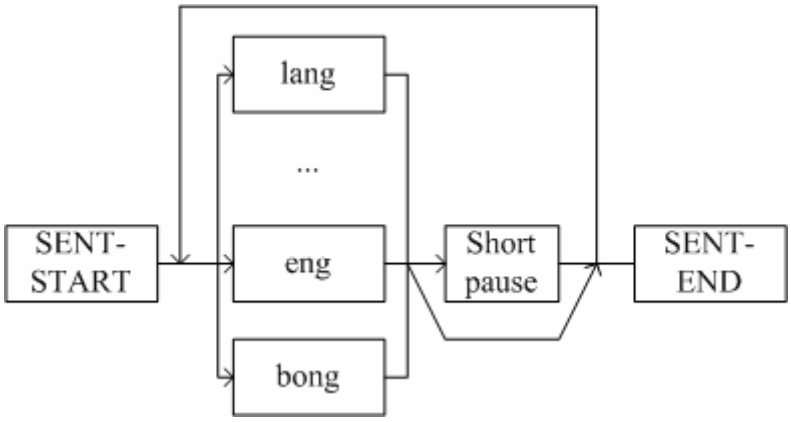

圖 2-4 辨識網路

2.8.分析辨識結果

由於我們的台語語料並不充足，因此我們將絕大部份的語料都用來作為訓練語料，僅拿 Database3 的少部份來當作 outside test 的辨識語料其總字數為 900 個字，表 2-4 為 outside test 之辨識率。

表 2-4 辨識率(outside test)

	音節總數(N)	Sub	Del	Ins	Recognition rate(%)
男	900	468	27	21	42.67
女	900	445	16	13	47.33

其中 $\text{Recong rate}=\frac{\text{N-(Sub+Del+Ins)}}{\text{N}}\times 100\%$ (2.1)

由上面的結果我們可以了解台語辨識率低於國語的辨識率，底下我們將針對這個現象提出了二個可能的原因。

1. 台語的基本音節數為 877 個，高於國語的基本音節數 411 個，相對地其聲學模型也比國語多，因此辨識錯誤的機率也相對的提高。
2. Stop ending (p，t，k，h) 的掌握並不容易，根據語言學家的研究指出，現代人在說台語時，對於 stop ending 的掌握愈來愈不明確，此一現象我們可由 confusion matrix 觀察得知。

首先，我們將韻母分成兩類，一類為基本韻母，一類為入聲韻母(p,t,h,k)，

C1={a,e,i,ia,io,iu,m,ng,o,oa,oe,ou,u,ui},

C2={ah,ak,ap,at,eh,ek,ih,ip,it,iah,iak,iap,iat,ioh,iok,iuh,mh,ngh,oh,ok,op,oah,oat,oeh,ouh,uh,ut,uih}，

並利用 confusion matrix 觀察其相互辨識關係，底下為其結果。

表 2-5 入聲韻母與基本韻母之相互辨識關係一

	C1 (%)	C2 (%)	Others (%)
C1	68	17.04	14.93
C2	23.4	48.14	28.46

由此我們可以看出類別 C2 較容易辨識為類別 C1，也就是說有 stop ending 的模型較易辨識為無 stop ending 的模型，同時也指出 C1 對 C2 或 C2 對 C1 的影響皆與 Others 對兩者的影響相差不多 (others 為不屬於這兩類的其他模型)。

接下來，我們將 C1 與 C2 細分，並重新配對，{{a} {ah,ak,ap,at}}，{{e} {eh,ek}} ， {{i}{ih,ip,it}} ， {{ia}{iah,iak,iap,iat}} ， {{io}{ioh,iok}} ，{{iu}{iuh}}，{{m}{mh}}，{{ng}{ngh}}，{{o}{oh,ok,op}}，{{oa}{oah,oat}}，{{oe}{oeh}}，{{ou}{ouh}}，{{u}{uh,ut}}和{{ui}{uih}}，其結果見表 2-6：

各欄位說明如後：Ref：辨識結果之參考答案；HIT：正確率；S(class)：入聲韻母辨識為相對應的基本韻母所佔之比例；S(other)：入聲韻母辨識為不屬於本身與相對應的基本韻母所佔之比例；N：參考答案出現的總次數。

表 2-6 入聲韻母與基本韻母之相互辨識關係二

Ref	HIT(%)	S(class)(%)	S(other)(%)	N
ah,ak,ap,at	31.7	10.96	57.32	82
eh,ek	45.45	45.45	9.1	44
iah,iak,iap,iat	27.9	6.98	65.11	43
ih,ip,it	62.22	16.67	21.11	90

ioh,iok	69.7	3.03	27.27	66
oah,oat	38.89	44.44	16.67	18
oh,ok,op	34.78	26.087	39.13	23
uh,ut	60	30	10	10

由此結果可以得知類別 C2 裡的韻母較易辨識為基本韻母為{eh,ek}，{oah,oat}，{uh,ut}的音節分別較易辨識為{e}，{oa}和{u}。

第三章 台語文句分析

3.1.簡介

由於國語文章並不能完全取代台語的意思，因此希望直接以台語文章來當合成器的文章。不過台語文章的寫法並不像國語文章一樣完全由漢字所組成，目前台語文章的寫法主要可分為三類：全拼音、全漢字和漢羅並用，例如：“au pai”、“後擺”和“後 pai”都是代表著國語“下一次”的意思，也就是說台語文章在寫作時對於拼音和漢字的選用並無統一的規定，目前常用的拼音系統包括：教會羅馬(本實驗室所用之系統)、通用拼音和教育部公告台灣閩南語的音標系統簡稱 TLPA。因此同一個詞讓不同的人來寫，或即使讓相同的人來寫也會產生許多不同的寫法，例如：“無涼”和“無當”都代表著國語“不行”的意思，它們的唸法也相同，即使詞典有收錄此詞也會因為寫法的不同而無法正確斷詞，因而降低了斷詞器的準確性。本論文將探討此問題。

3.2.基本架構

文字分析主要由下列幾個部份所組成：文字正規化(Text normalization)、斷詞、構詞和詞轉音【7】【8】：

1. 文字正規化：在書寫文章時由於某些詞並不會以口語化的方式書寫，因此在本模組裡，主要將文章裡不口語化的寫法部分轉換為較口語化的寫法。

例如：

寫法：1445 公斤	讀法：一千四百四十五公斤
寫法：45%	讀法：百分之四十五
寫法：12:50	讀法：十二點五十分

目前我們只處理第一種寫法，對於特殊符號與小數點的部份還未處理。

2. 斷詞：主要是找出一段話裡每個詞的邊界，此單元包含二個程序：建立格狀詞組和動態規劃詞序列搜尋，底下為斷詞器的實現步驟。

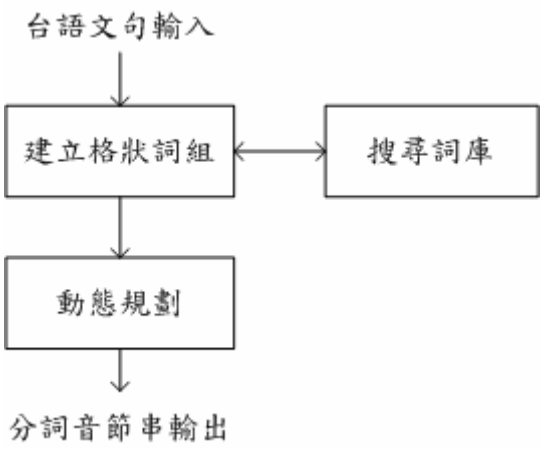

圖 3-1 斷詞器架構

格狀詞組：我們以向前(forward)成詞的方法來建立詞格，如果所形成的詞格大於四字詞的話，由於詞格大於四字詞的並不多見，因此認為其錯誤率也最小，固相信詞格為四字詞以上的詞，再以此最大詞長為結束點，並以此詞的下一個字當起始點再繼續建立詞格。以圖 3-2 為例，以「高」起頭的詞有「高」、「高雄」和「高雄車站」，因已找到四字詞「高雄車站」，所以就從「欲」繼續建立詞格。而「雄」、「車」、「站」和「車站」，就不形成詞格。

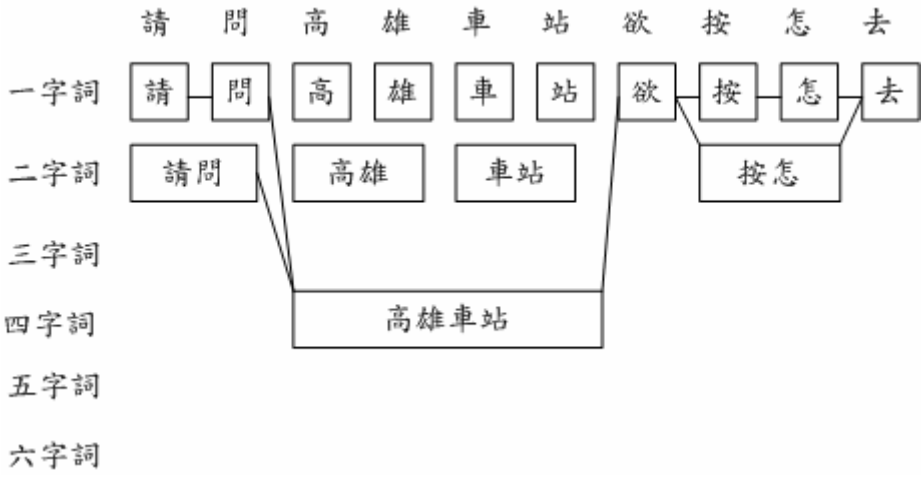

圖 3-2 格狀詞組範例

動態規劃：建立好的每個詞格都視為一個節點，並給每個節點一個分數，其中每個節點皆可與前後相鄰之節點相連結，將路徑上的每個節點之分數相加，即為此路徑的最終分數。在此我們所用的斷詞規則為長詞優先，底下為計分方式：

$$S(W)=\sum_{i=1}^{N} s(w_i), \qquad \text{其中} \quad s(w_i)=\begin{cases} 1, \text{ if } L(w_i)=1 \\ 4, \text{ if } L(w_i)=2 \\ 6, \text{ if } L(w_i)=3 \\ 16, \text{ if } L(w_i)=4 \\ 25, \text{ if } L(w_i)=5 \\ 36, \text{ if } L(w_i)=6 \end{cases} \tag{3.1}$$

$W=(w_1,w_2,...,w_n)$ 為待選詞串，$L(w_i)$ 為詞 w_i 之詞長

上列之計分方式主要是要讓二字詞接二字詞的組合優於一字詞接三字詞或三字詞接一字詞的組合，並依據維特比搜尋(Viterbi search)找出分數最高之路徑，即為所要的詞串輸出。

3. 構詞：因詞典無法列舉出所有的詞，因此需利用構詞規則，將詞構出，此部份需要每個斷完詞後的詞性，再依據合法的詞性組將詞構出。但因目前台語詞典裡對於詞性的資訊尚不齊全，故此部份只將疊詞合併。

4. 詞轉音：因為台語常會有一詞多音的現象，故我們將斷詞後的結果標上較常發的音，再經由變調規則轉換成相對應的羅馬拼音。

3.3.問題描述

由於台語文章主要是由漢字與拼音所組成的，底下我們將台語文字寫法的問題分成拼音與漢字兩個部份來進行討論。

3.3.1. 拼音部份

在拼音部份，目前所遭遇的問題有二：(1) 拼音系統使用的問題；(2) 書寫方式的問題，底下將依據此二點進行討論。

(1) 各家所使用的系統並不統一，因此在斷詞時常因此造成斷詞上的錯誤，目前實驗室所收集到的拼音系統有教羅，通用以及 TLPA 教育部所公佈的拼音系統。

(2) 在拼音的寫法方面也各有不同，例如：“e-tang” 和 “etang” ，前者有一個分離符號將音節分開，因此不會誤判為不合法的詞，後者因無法輕易地將基本音節分開因此可能被判定為不合法的詞或為一字詞。

3.3.2. 漢字部份

漢字的選用為何常會受個人用字的習慣有關，因而造成使用漢字不一的問題，例如：”台灣人”和”台灣儂”都代表著國語台灣人的意思，但是詞典並不可能將所有的寫法都收錄，因而降低了斷詞器的正確率，例如：”台灣儂”以目前的斷詞器將會斷成”台灣/儂”，底下將針對漢字的使用情形作討論。

使用較固定的漢字詞多數屬於下列兩種情形(其中第一點並不是絕對)：

(1)來源自韵書和古文裡的詞：可在古文或韵書中找到音義皆適合的漢字詞。代表這種詞的漢字其意義和發音皆為固定，其中他們的發音分為兩類：文言音和白話音。(註：韵書為古時候的詞典，內容皆為漢字)。

(2)表示同一意義並且台語和國語用同根詞：例如： “family”、“political” 等意義的國語和台語都用同根詞，「家庭」、「政治」。而使用同語根的詞多半是實詞(content 或 lexical words，包括名詞、動詞和形容詞)，少見於虛詞(grammatical functional 或 empty words，包括介詞、副詞、聯結詞、量詞、感嘆詞、助動詞等)，台語用字的分析見表 3-1【9】。(註：破音字大多屬於同根詞，例：睡覺的覺，覺得的覺，則屬同根詞)。

而漢字使用不固定則屬下面相反的情況:代表同一意義但是台語和國語卻使用不同語根的詞,例如:“eye”發台語音用 bak-chiu,漢字寫法為『目珠』或『目睭』,國語用『眼睛』;“at last”發台語音用 soah-boe 漢字寫法可寫成『煞尾』或『息尾』 國語用『最後』。而在使用不同語根之情況下,虛詞所佔的比例,高於實詞。

表 3-1 台語用字的分析

	任兩個作者文章之間使用同字的機率	任兩本辭典之間使用同字的機率	與國語用同語根的機率
名詞及轉用動詞形容詞(革命、動員、科學化)	90%	95%	80
動詞(欣賞、看)	70%	80%	60%
形容詞(見笑、臭)	70%	80%	60%
連結詞(對、到)	40%	50%	10%
助動詞(會、beh)	50%	60%	30%
感嘆詞、語氣詞(呢、哩、啊)	50%	60%	40%
代名詞、指示詞、疑問詞(你、che、啥人)	70%	70%	40%
複音詞(革命、垃圾)	92%	95%	85%
單音詞(人、紙、走)	85%	90%	60%

底下依據不同的觀點來分析台語的用字。

1.根據詞的表示法,我們可以將台語詞的表示法,分成字音皆同,字異音同和易造成混淆的字三類。底下舉例說明:

表 3-2 字音皆同

拼音	漢字
kok-ka	國家
tai5-oan5	臺灣

choa3	紙
ian5-pit8	鉛筆
sin-bun5	新聞

表 3-3 字異音同

拼音	常用的漢字
li2	你、汝
chit	這、此
goan2	阮、我
m7	不、唔
他 long2 無來	都、攏
一 chang5 樹	棵、欉

表 3-4 易造成混淆的字

例句	常用的漢字
我 ka7 他拍	給、供、把、將
kha-chhng	尻川、尻穿、腳穿
si3-koe3[ke3]	四界、四屆
真 che7 人	濟、多
線 chhoah 斷	擦、掇
chhap 人的家內事	插、管、採

在易造成混淆的字之情況下，常會因字的發音不同而產生不同的意思，例如：『我給(ka)他拍』和『我給(hou)他拍』，所用漢字皆一樣，但所代表意思卻不同，前者之國語義為『我打他』，後者之國語意為『他打我』。

2.在上下文中對於相同之漢字是否有相同之使用方法, 依據此方面的探討，我們可以把台語裡的常用漢字分成音義固定、音異而意義固定和音義皆不固定等三類。底下舉例說明：

表 3-5 音義都固定的漢字

例詞	拼音
戰爭	chian3-cheng
世界	se3-kai3
時陣	si5-chun7
應當	eng3-tong
電視	tian7-si7
布袋戲	pou3-te7-hi3

表 3-6 發音不固定而意義固定的漢字

例詞	常用的拼音
大學	toa7-oh, tai7-hak8
東南亞	tang-lam5-a, tong-lam5-a
天使	tiN-sai3, tian-sai3
讚美	o2-lo2, chan3-bi3
幫助	pang-chan7, pang-chou7

表 3-7 意義發音都可能產生誤會的漢字

例詞	不同拼音與相對應之意義
伊	i 他 in 他們
藏	khng3 放 chhang3 藏
給我拍	hou7 給、讓、被、 ka7 把、給
您	li2 您 lin2 你們

對於意義、發音都可能產生誤會的漢字，我們可由上下文的關係推測出字的念法與意義，例如：『有一個人，伊(i)問我火車站在哪』、『有一陣人，伊(in)問我火車站在哪』，前者裡的『伊』代表『他』的意思，後者『伊』代表『他們』的意思。可是我們卻可從『一個人』與『一陣人』來判定『伊』的念法與意義。

3.一般人選字的習慣：

(1)假借：因國語字的台語發音與台語詞的發音相同或相近因而被借去代表台語詞。但是這種情形並不多，可是卻是比訓用字容易唸出正確的發音，舉例如下：

表 3-8 假借

拼音	相漢字	例詞	國語的意思
chhng	穿	尻穿	屁股
tau	鬪	柴鬪柄	裝
cha-bou	查某		女人
chhai	採	無採	可惜

(2)轉用漢字：因國語詞的音義皆近而轉用漢字，一個詞雖然找不到發音相同的國語字，但卻可以選用發音相近、意義又相近的字來代表。舉例如下：

表 3-9 轉用漢字

臺語詞	轉用漢字	該字文言或白話音
khah	較	「較好 khah ho」 「比較 pi-kau」
That	塞	「路塞哩 lou that-leh」 「塞嘴齒 she chhui-khi」
Chhoa	導	「導路 chhoa lou」 「指導 chi-to」
Kham	蓋	「蓋蓋 kham koa」
Thau	釋	「釋索仔 thau soh-a」 「解釋 kai-sek」
Bauh	貿	「貿貨 bauh hoe」 「貿易 bou-ek」
Theh	提	「提物 theh mih」 「提議 the-gi」
hoaN	按	「按站壁 hoaN tiam piah」 「按算 an-sng」

(1) 訓用漢字(同義字)：借用古漢語或國語文裡的意義(包括語法意義)和台語詞的意義相似。這種漢字的用法，稱為訓用漢字。因國語裡的意義而使用此字的，稱為國語訓；因古文裡的意義而使用的，稱為古文訓，底下舉例說明：

表 3-10 古文訓的例子

臺語詞	訓用韵字	其他通行字
phaiN	惡	歹
Sui	美	
Beh	欲	卜、要
ia-be	尚未	猶未
Bo	妻	某
siu-khi	怒氣	受氣

表 3-11 國語訓的例子

拼音	訓用字	例詞	其他常用字
She	著	倚著看	的、地
She	繞	繞到彼旁	旋
Gia	抬		迎
Khia	堅		倚、企
Tak	各	各項	逐項

3.4.詞典的合併與處理

為了要降低台語文字寫法混亂進而影響台語斷詞正確的問題，我們將合併數本詞典，並利用合併後的詞典來產生字對音與音對字的表，進而來降低因台語文字不統一對斷詞器所造成的影響。

目前我們所擁有的資料如下：(1) 實驗室本身所擁有的詞典共 110,797 筆詞條；(2) 從“一分鐘台語單字速成”中選常用的詞共 1675 筆詞條，其資訊包含台語漢羅、台語拼音、相對華文；(3) 鄭良偉教授所提供的詞典共 52,209 筆詞條，其資訊包含台語漢羅、台語拼音、相對華文、詞類(依中央研究院定義)和華語的

詞頻。由於“一分鐘台語單字速成”中所用的拼音系統為 Taiwan Language Phonetic Alphabet (TLPA)，與實驗室所用的教會羅馬拼音系統不同，因此先將拼音轉換成實驗室所用的拼音系統。對於非實驗室的詞典我們作以下的處理：刪除重複的詞(此處所說重複的詞是指台文寫法與拼音皆與實驗室相同)、剔除非教會羅馬拼音的詞和挑出台文與拼音無一對一關係的詞，整理後共增加 3141 筆詞條，資料統計如下：

表 3-12 詞典資料統計

	詞條數
一字詞	21042
二字詞	69291
三字詞	18102
四字詞	5170
五字詞	317
六字詞	17
總數	113939

再利用合併後的詞典產生字對音和音對字的表，並記錄相對的音(字)在詞典中總共出現的次數和分別在詞中的各個位置出現的次數，其格式如下：

表 3-13 字對音格式說明

字	音一	此音在詞典中出現的次數	在詞中位置出現的次數	例詞
口	kau	3	[0 2 1 0 0 0]	進人口

表 3-14 音對字格式說明

音	字一	此字在詞典中出現的次數	在每個位置出現的次數	例詞
Ap	壓	60	[18 24 8 0 0 0]	變壓器

我們可應用這兩個表在詞轉音的階段，根據每個字在詞中的位置將其音標上。在

辨識時，我們可根據這兩個表將辨識所得的音轉換為相對應的字。

3.5.處理拼音符號連在一起寫的問題

所謂拼音符號連在一起的意思是說，連續的拼音符號沒有用分隔符號區分開來，例如：etang，如果有用分隔符號則為 e-tang，上面兩個例子都代表華語“可以”的意思，但前者可能會被判定為一字詞，或不合法的詞，我們將此問題視為和斷詞一樣，我們稱之為斷音，此步驟包含兩個程序：建立格狀音組和動態規劃音序列搜尋，底下為斷音器的實現步驟：

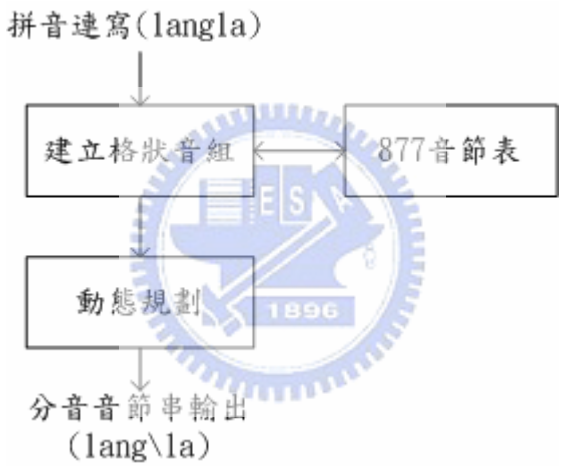

圖 3-3 斷音器架構

格狀音組：在此我們用了建立格狀音組的方法與斷詞的方法一樣，不同的是並不是所有單獨的英文字母都是合法的基本音節，以圖 3-4 為例「p」並不是一個合法的基本音節，因此不形成音格；在教會羅馬拼音系統裡，最長的基本音節是為 8 個英文字母。

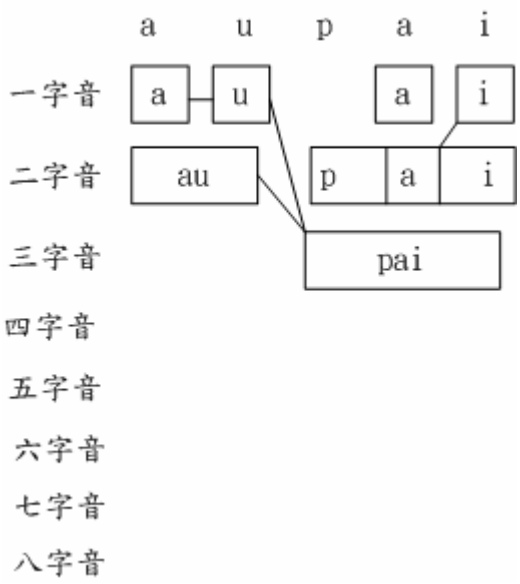

圖 3-4 格狀音組範例

動態規劃：建立好的每個音格都視為一個節點，並給每個節點一個分數，其中每個節點皆可與前後相鄰之節點相連結，累加每條路徑上各個節點的分數，並依據 viterbi search 找出最佳之路徑，決定出音的邊界，在此我們所用的斷音規則為英文字母最多的音為優先，底下為計分方式：

$$S(W)=\sum_{i=1}^{N}L^2(w_i) \qquad (3.2)$$

其中 $W=(w_1,w_2,...,w_n)$ 為待選音串，$L(w_i)$ 為音 w_i 之字母個數。

3.5.1. 實驗步驟

由於目前我們並無大量屬於此類的文章，因此我們利用詞典來當做測試資料，我們的作法是先將詞典裡的羅馬拼音取出並將分隔符號與聲調去除，例如：ka7-mng7 我們將分隔符號與聲調去除後即為 kamng。底下為我們測試出來的結果。

表 3-15 斷音器之實驗結果

總測試詞條(N)	正確詞條(H)	錯誤詞條(E)	正確率($\frac{H}{N}$×100%)
74150	67299	6851	90.76%

利用類似長詞優先的規則雖然可解決大部份的問題，但還是有些問題尚未解決，例如：”si”與”an”連寫成”sian”，由於”sian”是屬於基本音節因此會造成錯誤，往後可利用語言模型來當給分的標準。

3.6.寫法和詞典不同的詞

寫法與詞典不同的詞可分為以下幾類：(1) 詞典是為漢羅並用的寫法，但文章卻是全漢字或全羅馬拼音的寫法；(2) 詞典為全漢字的寫法，文章是漢羅並用的寫法；(3) 詞典為全羅馬拼音的寫法，文章是為全漢字或漢羅並用的寫法；(4) 詞典中漢字的寫法，與文章中漢字的寫法不同。我們將各類改善之結果置於 3.6.2 節，表 3-16。

首先我們先針對第一、二類即詞典是漢羅並用的寫法，但文章卻是全漢字或全羅馬拼音的寫法與詞典為全漢字的寫法，文章是漢羅並用的寫法的問題做處理，其處理流程見圖 3-5。

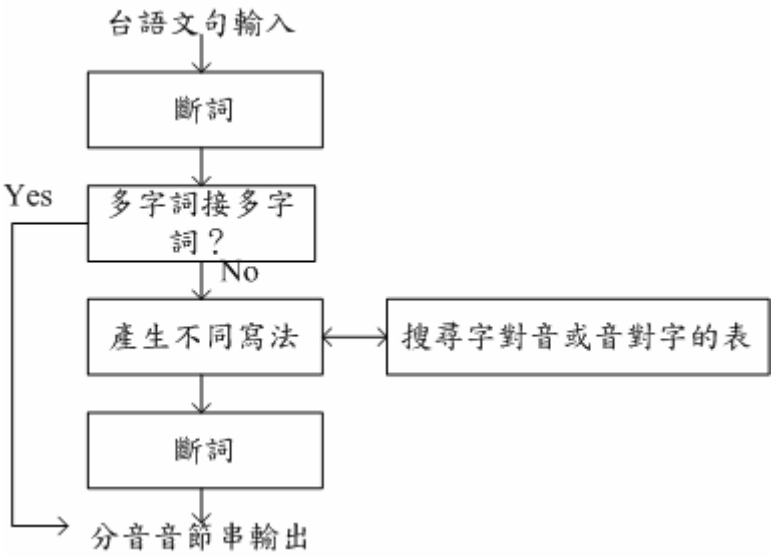

圖 3-5 表示法混亂之處理流程

上述流程圖裡的文句分析和斷詞單元，已在前面說明過，底下我們將藉由實例來說明產生不同寫法方塊的動作：

例句：這有一封 phoe 信

斷詞後：這有\一封\phoe\信

在這個例句中“phoe 信”應該要被斷成為二字詞，不過由於我們的詞典沒收錄此種寫法，因此會誤判為兩個一字詞“phoe”及“信”，因此將這兩個連續的一字詞收集起來，再將二字詞中所有以“phoe”開頭和以“信”結尾的二字詞全都列出，並在所列出的詞中將不符合寫法的字，利用字對音或音對字的表加以替換，以期望找到符合輸入的詞。例如：輸入的是“phoe 信”一詞，詞典找不到，列出 beh 信、ka 信、m 信、十信、批信、私信…等。

利用字對音的表將所列出的詞的第一個字以音替換，並和輸入的詞比較是否一樣，一樣即暫定為二字詞 (在此例，由“批信”一詞中，我們可找到“phoe 信”的組合，因為“批” = “ phe” = “phoe”)，再進行斷詞的動作。這樣對我們來說可以不用對每個詞做展開的動作，因為我們所要展開詞的條件為：詞中的字其寫法要和輸入的詞一樣並且位置也必須一樣。如此一來可拿來展開的詞數就相對的少很多了。

接下來我們將針對第三類與第四類的問題作處理，此部份的處理方式只是處理第一類問題的延伸；在此部份裡我們將字對音和音對字的轉換共作了兩次，也就是說，將符合第一個字或第二個字的詞全部列出後，將詞中不符合的字或音轉換為音或字時，再一次的將轉換後的音或字轉換為字或音，才與輸入的詞作比較，相同則暫定為詞，再進行斷詞的動作。

接下來我們將說明針對三字詞(含)以上的處理方式，也就是說原本是三字詞(含)以上的詞但因寫法與詞典不同而斷成連續的一字詞，或多字詞接一字詞，例如：”台灣儂”，因儂的寫法與詞典不同而斷成”台灣 儂”，處理的方法大致與處理二字詞一樣，不同的是三字詞以上的詞是固定二字詞(含)以上的詞只針對一字詞作展開的動作再與輸入的詞作比較，主要的原因是因為要是針對連續的一字詞做展開的動作的話較沒效率而且速度也會非常慢，正確率也無明顯的改善，所以目前我們只針對多字詞接一字詞或一字詞接多字詞的情形做處理，往後再用較有效率的方法解決連續被斷成一字詞的多字詞，也許可利用上述之作法一步步地將之合為多字詞。

3.6.1.實驗步驟

1. 收集台語文章

期望能收集到的台語文章其總字數約在 10 萬至 100 萬字之間，並盡可能的涵蓋各區域及各層面的文章，使其具有代表性及豐富性。

文章來源：

- 鄭良偉教授現有的文章電子檔
- 社會上各個鄉土研究團體所發表的書籍、期刊
- 網路上所發表的文章

收集之台語文章的書寫格式：

- 全漢字書寫
- 漢羅並用書寫

2. 文章前處理

i. 漢字部份：

➢ 自創字的處理

我們可將字轉成相對應的拼音，在此步驟我們需要的資訊為：作者是為了何音而造出此字，還有此字對應到 windows 內碼為何，以及此字對於其他作者而言是否有不一樣的用途與不一樣的發音，底下我們列出網路上

常用的造字，其字型檔為 Taiwanese Serif for Windows 或 Taiwanese Fixed for Windows。詳見附錄二

- 組合字

組合字的存在，例如：“勿會”組合成  ，目前我們是將之當成兩個字來進行處理。

在上述的各種情形，除了 Taiwan Package 可以自動轉換外，其餘皆須仰賴人工之處理，因此若數量不是很多的話，或許我們可以暫且忽略該詞。

ii. 拼音部份：

- 拼音系統轉換

目前的台語拼音系統尚未一統，因此在漢羅並用的文章裡，我們可能會面對到拼音系統的不同問題。因此在這部分，我們需解決之問題有：

1. 建立台語拼音系統互換對照表，目前收集到的拼音系統為教會羅馬拼音，通用拼音和 TLPA。
2. 將文章中之拼音轉換為教會羅馬拼音

- 移除調號

目前台語音調部分也尚無共同之標準，調號所代表之真正聲調皆因人而異。另外，我們不知文章中標的調為變調後之調或本調，更無法檢查其是否聲調打錯。因此，我們將把文章中羅馬字有標上調號的部分，將其調號移除。除此之外，我們尚會遇到一些特殊的調號標示方式，如：非以阿拉伯數字置於羅馬字後的調號表示方式，如：

1. “pak-hôe-sòa°”
2. “ ke^{7}”
3. 厂乇 3 sing3 (蒼蠅)

上述情形需另加以轉換。

iii. 大小寫及全半形之轉換：

在文章中的數字，我們皆須將其從全形轉換為半形。而英文字母的部分，除了需將其從全形轉換為半形之外，另需考慮除“N”外（因為在我們的拼音系統中，N 和 n 是有分別的），其餘需將大寫轉換為小寫。

iv. 特殊符號的使用：

1. an-ne 分環境，符號“-”將音連接起來當成一個詞

2. 無排練過是無法度表演 =e，符號“=”或許代表 e 該發輕聲調。

此部分，我們將“-”當成音節和音節邊界，對於“=”目前的是將它捨棄。

3.6.2.實驗分析與其意義

實驗的目的，主要是要對一字詞接多字詞與多字詞接一字詞的詞再進一步的處理，期望能構成有意義的詞，進而降低因寫法的混亂所造成的影響，底下我們將介紹實驗結果。

我們將以三種不同的角度來觀察我們處理過後的結果：(1) 處理過後所得的詞其意義和唸法皆與詞典相同；(2) 處理過後所得的詞其意義與詞典相同但唸法不同；(3) 經人判斷處理過後的詞是為一個合理的詞，即便意義與唸法皆與詞典不同，並根據 3.6 節的分類觀察各類別所佔的比率，與各類別的結果。

表 3-16 為各類別經處理後所佔之比例：

表 3-16 各類別所佔之比例

第一類(%)	第二類(%)	第三類(%)	第四類(%)	處理後所得到的詞數
19.56	4.695	0.116	75.59	2576

由表 3-16 中我們可看出二、三類所造成的問題比例較一、四類低，可能原因有二：(1) 當詞典的詞是以全漢或全羅的方式表示時，人們會以漢羅或全羅來表示同一個詞的機率並不高。(2) 我們所用之方法尚無法解決此二類的問題。由於一、四類佔的比例較高，因此接下來我們觀察此作法對一、四類改善的程度。

表 3-17 第一類的結果

音意相同(%)	意同(%)	音意不同(%)	處理後所得之詞數
87.3	88.3	93.85	504

表 3-18 第四類的結果

音意相同(%)	意同(%)	音意不同(%)	處理後所得之詞數
27.75	28.73	63.05	1946

各欄位說明如後：音意相同：斷出來的詞其發音與意義皆與詞典一樣；意同：發音與詞典不同，但意義卻相同；音意不同：經人判斷處理過後的詞是為一個合理的詞，但其音和意卻與詞典不同。

由表 3-17 中可知當寫法不一的情形是屬於第一類的話，我們利用此種作法可得到較不錯的結果，不論是針對詞的意思或詞的標音；再觀表 3-18 的結果當寫法是屬於第四類的話，根據音意相同而辨定為詞的比率並不高，但如果在不考慮音、意皆要相同的情況下，能辨定是為詞的比率有六成，之後可再經由詞轉音的單元將音標示上，

經由以上的實驗後，我們發覺還有其他因素會影斷詞正確性，底下我們將問題一一列出：(1) 詞典中收錄了不適當的詞，例如：講台，雖然是適當的國語講法，但在台語裡卻是不適合的講法；bat 聽，在台語中並不足以當詞，bat 聽過才是詞，對於此類的詞應從詞典中將之剔除；(2) 在斷詞結果中，有此常用詞並沒有斷出，例如：警察局，經檢查後發現是詞典無收錄此類的詞，因此對於常用詞的收集還有待加強；(3) 有些作者在寫作文章時會將不易念的字將其音標上例如：匏(pu)仔，在往後的斷詞時應以其中一個為依據來進行斷詞；(4) 由於第四類佔了問題的絕大部分，但其辨定為詞的比例卻不高，因此當我們利用字對音和音對字的表來進行替換的動作時因以出現次數較多的字(音)來進行替換的動作；(5) 在構詞時，我們可將仔、兮與前面的詞合併，例如：斷出來的詞為”真大/兮”經合併後為”真大兮”，如此的說法較為口語化。

第四章 結論與未來展望

4.1.結論

本論文包含台語語音辨識與台語文句分析，在台語語音辨識部分，由於語料收集不易因此我們只求將台語的聲學模型補齊，並利用所訓練而來的聲學模型進行台語單音節語音辨識，其男女辨識率分別為 42.67%與 47.33%，雖然拿來做測試的語料並不多，不過我們卻可從辨識結果中確定入聲韻母對辨識系統的影響，其中較易辨識成基本韻母的入聲韻母是{eh,ek}，{oah,oat}，{uh,ut}分別較易辨識為基本韻母{e}，{oa}和{u}，因此對於此類的入聲韻母可考慮將之合併於相對應的基本韻母，以期提高辨識率。

在台語文句分析部份，之前並無考慮到拼音連寫與台文表示法不一的問題例如：etang = e tang，在本論文裡我們採取英文字母最長的拼音為優先的規則製做了斷音器主要是為了要降低拼音連寫對文句分析單元所造成的問題，並以詞典為測試資料，其斷音正確率為 90.76%，而在台文表示法不一的問題裡，我們藉由對輸入的詞做有條件式的擴充，以期降低台文表示法混亂的問題，並分類觀察不同的結果，得知，此法可降低一、四類的寫法問題，並且如果寫法是屬於第一類的話，即詞典是為漢羅並用的寫法，但文章卻是全漢字或全羅馬拼音的寫法，其音義皆相同所佔的比例高於其他三類。

4.2.未來展望

在台語文句分析單元裡尚有些問題有待討論，與未來需改進的地方。

1. 將詞典以樹狀結構的方式存之，在往後增加詞條時可以較有效率的方式存之，並且不會因為詞典變大而使搜尋典詞的速度變慢。
2. 目前的斷音器，我們所針對的拼音系統只限於教會羅馬拼音系統，對於其他則尚未考慮，因此可增強斷音器的功能使其能處理各種拼音系統。

3. 在觀察降低台文表示法混亂的結果時，有一類佔了相當重的比例，即意義和拼音皆和辭典不同，但經人判斷是可當成詞，此類如能將其音正確標上，相信對台語合成器能有一定的幫助。
4. 在文字正規化的部份，目前對於特殊符號的表示法還未處理例如：$45.6。
5. 目前斷詞(音)器的給分方式是以詞(音)長為主給定一個分數，並不公平，往後可利用語言模型做為斷詞(音)器的給分標準，以增加斷詞(音)的正確率。
6. 在詞典方面，對詞的選用應再嚴格一點，盡量選用台語常用的詞，不要直接利用國語的詞，例如講台，台語中並無此種說法。
7. 將經常使用的字加入詞典中，例如：”儂”，經常與”人”通用，因此可將之加入詞中，例如：詞典中有”台灣人”可再將”台灣儂”加入。

參考文獻

【1】 陳佩玥，“台灣閩南語中手部動作特指「打」的語義探析”，第五屆漢語詞彙語義學研討會論文集，新加坡，2004年6月。

【2】 王閔鴻，“不特定語者大辭彙華台雙語辨識引擎之研製及其應用”，私立長庚大學碩士論文，民國九十二年六月。

【3】 鄭良偉，”臺語、華語的結構及動向”，遠流出版社，1997年。

【4】 方南強，”臺語通用會話：一套讓你能說出台語美辭雅語的教材.第二集”，開拓出版商，1997年。

【5】 鄭如玲，”一分鐘台語單字速成”，三思堂文化事業有限公司，2002年。

【6】 S. Young, G.. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. Woodland, “The HTK Book（for HTK Version 3.2.1）”

【7】 楊鈺清，”台語文句翻語音系統之製作”，國立交通大學碩士論文，民國八十八年六月。

【8】 傅振宏，”基於自動產生合成單元之台語語音合成系統”，私立長庚大學碩士論文，民國八十九年六月。

【9】 鄭良偉，”走向標準化的台灣話文”，自立晚報出版社，1989年。

附錄一

音碼	音節	聲母(119類)	韻母(84類)
1	a	INULL_2	a
2	ah	INULL_2	ah
3	ai	INULL_2	ai
4	aiN	INULL_2	aiN
5	ak	INULL_2	ak
6	am	INULL_2	am
7	aN	INULL_2	aN
8	an	INULL_2	an
9	ang	INULL_2	ang
10	ap	INULL_2	ap
11	at	INULL_2	at
12	au	INULL_2	au
13	ba	b_2	a
14	bah	b_2	ah
15	bai	b_2	ai
16	bak	b_2	ak
17	ban	b_2	an
18	bang	b_2	ang
19	bat	b_2	at
20	bau	b_2	au
21	bauh	b_2	auh
22	be	b_3	e
23	beh	b_3	eh
24	bek	b_3	ek
25	beng	b_3	eng
26	bi	b_4	i
27	bian	b_4	ian
28	biat	b_4	iat
29	biau	b_4	iau
30	bih	b_4	ih
31	bin	b_4	in
32	bio	b_4	io
33	bit	b_4	it
34	biu	b_4	iu
35	bo	b_5	o
36	boa	b_5	oa
37	boah	b_5	oah
38	boan	b_5	oan
39	boat	b_5	oat
40	boe	b_5	oe
41	boeh	b_5	oeh
42	boh	b_5	oh
43	bok	b_5	ok
44	bong	b_5	ong
45	bou	b_6	ou
46	bu	b_7	u
47	bui	b_7	ui
48	bun	b_7	un
49	but	b_7	ut
50	cha	ch_2	a
51	chah	ch_2	ah
52	chai	ch_2	ai
53	chaiN	ch_2	aiN
54	chak	ch_2	ak
55	cham	ch_2	am
56	chan	ch_2	an
57	chaN	ch_2	aN
58	chang	ch_2	ang
59	chap	ch_2	ap
60	chat	ch_2	at
61	chau	ch_2	au
62	chauh	ch_2	auh
63	che	ch_3	e
64	cheh	ch_3	eh
65	chek	ch_3	ek
66	cheN	ch_3	eN
67	cheng	ch_3	eng
68	chha	chh_2	a
69	chhah	chh_2	ah
70	chhai	chh_2	ai
71	chhak	chh_2	ak
72	chham	chh_2	am
73	chhan	chh_2	an
74	chhang	chh_2	ang
75	chhap	chh_2	ap
76	chhat	chh_2	at
77	chhau	chh_2	au
78	chhe	chh_3	e
79	chheh	chh_3	eh
80	chhek	chh_3	ek
81	chheN	chh_3	eN

82	chheng	chh_3	eng
83	chhi	chh_4	i
84	chhia	chh_4	ia
85	chhiah	chh_4	iah
86	chhiak	chh_4	iak
87	chhiam	chh_4	iam
88	chhian	chh_4	ian
89	chhiaN	chh_4	iaN
90	chhiang	chh_4	iang
91	chhiap	chh_4	iap
92	chhiat	chh_4	iat
93	chhiau	chh_4	iau
94	chhih	chh_4	ih
95	chhim	chh_4	im
96	chhin	chh_4	in
97	chhiN	chh_4	iN
98	chhio	chh_4	io
99	chhioh	chh_4	ioh
100	chhiok	chh_4	iok
101	chhiong	chh_4	iong
102	chhip	chh_4	ip
103	chhit	chh_4	it
104	chhiu	chh_4	iu
105	chhiuN	chh_4	iuN
106	chhng	chh_8	ng
107	chhngh	chh_8	ngh
108	chho	chh_5	o
109	chhoa	chh_5	oa

110	chhoah	chh_5	oah
111	chhoaN	chh_5	oaN
112	chhoan	chh_5	oan
113	chhoe	chh_5	oe
114	chhoeh	chh_5	oeh
115	chhoh	chh_5	oh
116	chhok	chh_5	ok
117	chhong	chh_5	ong
118	chhou	chh_6	ou
119	chhu	chh_7	u
120	chhuh	chh_7	uh
121	chhui	chh_7	ui
122	chhun	chh_7	un
123	chhut	chh_7	ut
124	chi	ch_4	i
125	chia	ch_4	ia
126	chiah	ch_4	iah
127	chiam	ch_4	iam
128	chiaN	ch_4	iaN
129	chian	ch_4	ian
130	chiang	ch_4	iang
131	chiap	ch_4	iap
132	chiat	ch_4	iat
133	chiau	ch_4	iau
134	chiauh	ch_4	iauh
135	chih	ch_4	ih
136	chim	ch_4	im
137	chin	ch_4	in

138	chiN	ch_4	iN
139	chio	ch_4	io
140	chioh	ch_4	ioh
141	chiok	ch_4	iok
142	chiong	ch_4	iong
143	chip	ch_4	ip
144	chit	ch_4	it
145	chiu	ch_4	iu
146	chiuh	ch_4	iuh
147	chiuN	ch_4	iuN
148	chng	ch_8	ng
149	cho	ch_5	o
150	choa	ch_5	oa
151	choah	ch_5	oah
152	choai	ch_5	oai
153	choaiN	ch_5	oaiN
154	choaN	ch_5	oaN
155	choan	ch_5	oan
156	choat	ch_5	oat
157	choe	ch_5	oe
158	choeh	ch_5	oeh
159	choh	ch_5	oh
160	chok	ch_5	ok
161	chong	ch_5	ong
162	chou	ch_6	ou
163	chu	ch_7	u
164	chuh	ch_7	uh
165	chui	ch_7	ui

166	chuiN	ch_7	uiN
167	chun	ch_7	un
168	chut	ch_7	ut
169	e	INULL_3	e
170	eh	INULL_3	eh
171	ek	INULL_3	ek
172	eN	INULL_3	eN
173	eng	INULL_3	eng
174	ga	g_2	a
175	gai	g_2	ai
176	gak	g_2	ak
177	gam	g_2	am
178	gan	g_2	an
179	gang	g_2	ang
180	gap	g_2	ap
181	gau	g_2	au
182	ge	g_3	e
183	geh	g_3	eh
184	gek	g_3	ek
185	geng	g_3	eng
186	gi	g_4	i
187	gia	g_4	ia
188	giah	g_4	iah
189	giam	g_4	iam
190	gian	g_4	ian
191	giang	g_4	iang
192	giap	g_4	iap
193	giat	g_4	iat

194	giau	g_4	iau
195	giauh	g_4	iauh
196	gih	g_4	ih
197	gim	g_4	im
198	gin	g_4	in
199	gio	g_4	io
200	gioh	g_4	ioh
201	giok	g_4	iok
202	giong	g_4	iong
203	gip	g_4	ip
204	git	g_4	it
205	giu	g_4	iu
206	go	g_5	o
207	goa	g_5	oa
208	goan	g_5	oan
209	goat	g_5	oat
210	goe	g_5	oe
211	goeh	g_5	oeh
212	gok	g_5	ok
213	gong	g_5	ong
214	gou	g_6	ou
215	gu	g_7	u
216	gui	g_7	ui
217	gun	g_7	un
218	gut	g_7	ut
219	ha	h_2	a
220	hah	h_2	ah
221	hahN	h_2	ahN

222	hai	h_2	ai
223	haiN	h_2	aiN
224	hak	h_2	ak
225	ham	h_2	am
226	haN	h_2	aN
227	han	h_2	an
228	hang	h_2	ang
229	hap	h_2	ap
230	hat	h_2	at
231	hau	h_2	au
232	hauN	h_2	auN
233	he	h_3	e
234	heh	h_3	eh
235	hehN	h_3	ehN
236	hek	h_3	ek
237	heN	h_3	eN
238	heng	h_3	eng
239	hi	h_4	i
240	hia	h_4	ia
241	hiah	h_4	iah
242	hiahN	h_4	iahN
243	hiam	h_4	iam
244	hian	h_4	ian
245	hiaN	h_4	iaN
246	hiang	h_4	iang
247	hiap	h_4	iap
248	hiat	h_4	iat
249	hiau	h_4	iau

250	hiauh	h_4	iauh
251	him	h_4	im
252	hin	h_4	in
253	hiN	h_4	iN
254	hio	h_4	io
255	hioh	h_4	ioh
256	hiok	h_4	iok
257	hiong	h_4	iong
258	hip	h_4	ip
259	hit	h_4	it
260	hiu	h_4	iu
261	hiuN	h_4	iuN
262	hm	h_9	m
263	hmh	h_9	mh
264	hng	h_8	ng
265	hngh	h_8	ngh
266	ho	h_5	o
267	hoa	h_5	oa
268	hoah	h_5	oah
269	hoai	h_5	oai
270	hoaiN	h_5	oaiN
271	hoaN	h_5	oaN
272	hoan	h_5	oan
273	hoat	h_5	oat
274	hoe	h_5	oe
275	hoeh	h_5	oeh
276	hoh	h_5	oh
277	hok	h_5	ok

278	hoN	h_5	oN
279	hong	h_5	ong
280	hou	h_6	ou
281	houhN	h_6	ouhN
282	houN	h_6	ouN
283	hu	h_7	u
284	hui	h_7	ui
285	huih	h_7	uih
286	huiN	h_7	uiN
287	hun	h_7	un
288	hut	h_7	ut
289	i	INULL_4	i
290	ia	INULL_4	ia
291	iah	INULL_4	iah
292	iak	INULL_4	iak
293	iam	INULL_4	iam
294	iaN	INULL_4	iaN
295	ian	INULL_4	ian
296	iang	INULL_4	iang
297	iap	INULL_4	iap
298	iat	INULL_4	iat
299	iau	INULL_4	iau
300	iauN	INULL_4	iauN
301	ih	INULL_4	ih
302	ihN	INULL_4	ihN
303	im	INULL_4	im
304	in	INULL_4	in
305	iN	INULL_4	iN

306	io	INULL_4	io
307	ioh	INULL_4	ioh
308	iok	INULL_4	iok
309	iong	INULL_4	iong
310	ip	INULL_4	ip
311	it	INULL_4	it
312	iu	INULL_4	iu
313	iuN	INULL_4	iuN
314	ji	j_4	i
315	jia	j_4	ia
316	jiah	j_4	iah
317	jiak	j_4	iak
318	jiam	j_4	iam
319	jian	j_4	ian
320	jiang	j_4	iang
321	jiap	j_4	iap
322	jiat	j_4	iat
323	jiau	j_4	iau
324	jiauN	j_4	iauN
325	jih	j_4	ih
326	jim	j_4	im
327	jin	j_4	in
328	jio	j_4	io
329	jiok	j_4	iok
330	jiong	j_4	iong
331	jip	j_4	ip
332	jit	j_4	it
333	jiu	j_4	iu

334	joa	j_5	oa
335	joah	j_5	oah
336	joan	j_5	oan
337	joe	j_5	oe
338	ju	j_7	u
339	jui	j_7	ui
340	jun	j_7	un
341	ka	k_2	a
342	kah	k_2	ah
343	kai	k_2	ai
344	kaiN	k_2	aiN
345	kak	k_2	ak
346	kam	k_2	am
347	kaN	k_2	aN
348	kan	k_2	an
349	kang	k_2	ang
350	kap	k_2	ap
351	kat	k_2	at
352	kau	k_2	au
353	kauh	k_2	auh
354	ke	k_3	e
355	keh	k_3	eh
356	kehN	k_3	ehN
357	kek	k_3	ek
358	keN	k_3	eN
359	keng	k_3	eng
360	kha	kh_2	a
361	khah	kh_2	ah

362	khai	kh_2	ai
363	khaiN	kh_2	aiN
364	khak	kh_2	ak
365	kham	kh_2	am
366	khaN	kh_2	aN
367	khan	kh_2	an
368	khang	kh_2	ang
369	khap	kh_2	ap
370	khat	kh_2	at
371	khau	kh_2	au
372	khe	kh_3	e
373	kheh	kh_3	eh
374	khehN	kh_3	ehN
375	khek	kh_3	ek
376	kheN	kh_3	eN
377	kheng	kh_3	eng
378	khi	kh_4	i
379	khia	kh_4	ia
380	khiah	kh_4	iah
381	khiak	kh_4	iak
382	khiam	kh_4	iam
383	khian	kh_4	ian
384	khiaN	kh_4	iaN
385	khiang	kh_4	iang
386	khiap	kh_4	iap
387	khiat	kh_4	iat
388	khiau	kh_4	iau
389	khih	kh_4	ih

390	khim	kh_4	im
391	khin	kh_4	in
392	khiN	kh_4	iN
393	khio	kh_4	io
394	khioh	kh_4	ioh
395	khiok	kh_4	iok
396	khiong	kh_4	iong
397	khip	kh_4	ip
398	khit	kh_4	it
399	khiu	kh_4	iu
400	khiuN	kh_4	iuN
401	khng	kh_8	ng
402	kho	kh_5	o
403	khoa	kh_5	oa
404	khoah	kh_5	oah
405	khoai	kh_5	oai
406	khoaN	kh_5	oaN
407	khoan	kh_5	oan
408	khoat	kh_5	oat
409	khoe	kh_5	oe
410	khoeh	kh_5	oeh
411	khok	kh_5	ok
412	khong	kh_5	ong
413	khou	kh_6	ou
414	khu	kh_7	u
415	khuh	kh_7	uh
416	khui	kh_7	ui
417	khuiN	kh_7	uiN

418	khun	kh_7	un
419	khut	kh_7	ut
420	ki	k_4	i
421	kia	k_4	ia
422	kiah	k_4	iah
423	kiak	k_4	iak
424	kiam	k_4	iam
425	kiaN	k_4	iaN
426	kian	k_4	ian
427	kiang	k_4	iang
428	kiap	k_4	iap
429	kiat	k_4	iat
430	kiau	k_4	iau
431	kiauh	k_4	iauh
432	kim	k_4	im
433	kiN	k_4	iN
434	kin	k_4	in
435	kio	k_4	io
436	kioh	k_4	ioh
437	kiok	k_4	iok
438	kiong	k_4	iong
439	kip	k_4	ip
440	kit	k_4	it
441	kiu	k_4	iu
442	kiuh	k_4	iuh
443	kiuN	k_4	iuN
444	kng	k_8	ng
445	kngh	k_8	ngh

446	ko	k_5	o
447	koa	k_5	oa
448	koah	k_5	oah
449	koai	k_5	oai
450	koaiN	k_5	oaiN
451	koaN	k_5	oaN
452	koan	k_5	oan
453	koat	k_5	oat
454	koe	k_5	oe
455	koeh	k_5	oeh
456	koh	k_5	oh
457	kok	k_5	ok
458	kong	k_5	ong
459	kou	k_6	ou
460	kouN	k_6	ouN
461	ku	k_7	u
462	kui	k_7	ui
463	kuiN	k_7	uiN
464	kun	k_7	un
465	kut	k_7	ut
466	la	l_2	a
467	lah	l_2	ah
468	lai	l_2	ai
469	lak	l_2	ak
470	lam	l_2	am
471	lan	l_2	an
472	lang	l_2	ang
473	lap	l_2	ap

474	lat	l_2	at
475	lau	l_2	au
476	lauh	l_2	auh
477	le	l_3	e
478	leh	l_3	eh
479	lek	l_3	ek
480	leng	l_3	eng
481	li	l_4	i
482	liah	l_4	iah
483	liak	l_4	iak
484	liam	l_4	iam
485	lian	l_4	ian
486	liang	l_4	iang
487	liap	l_4	iap
488	liat	l_4	iat
489	liau	l_4	iau
490	lih	l_4	ih
491	lim	l_4	im
492	lin	l_4	in
493	lio	l_4	io
494	lioh	l_4	ioh
495	liok	l_4	iok
496	liong	l_4	iong
497	lip	l_4	ip
498	lit	l_4	it
499	liu	l_4	iu
500	liuh	l_4	iuh
501	lo	l_5	o

502	loa	l_5	oa
503	loah	l_5	oah
504	loan	l_5	oan
505	loat	l_5	oat
506	loe	l_5	oe
507	loeh	l_5	oeh
508	loh	l_5	oh
509	lok	l_5	ok
510	lom	l_5	om
511	long	l_5	ong
512	lop	l_5	op
513	lou	l_6	ou
514	lu	l_7	u
515	lui	l_7	ui
516	lun	l_7	un
517	lut	l_7	ut
518	m	m_1	FNULL
519	ma	m_2	a
520	mah	m_2	ah
521	mai	m_2	ai
522	mau	m_2	au
523	mauh	m_2	auh
524	me	m_3	e
525	meh	m_3	eh
526	mi	m_4	i
527	mia	m_4	ia
528	miau	m_4	iau
529	mih	m_4	ih

530	mng	m_8	ng
531	mo	m_5	o
532	moa	m_5	oa
533	moai	m_5	oai
534	moe	m_5	oe
535	moh	m_5	oh
536	mou	m_6	ou
537	mouh	m_6	ouh
538	mui	m_7	ui
539	na	n_2	a
540	nah	n_2	ah
541	nai	n_2	ai
542	naih	n_2	aih
543	nau	n_2	au
544	nauh	n_2	auh
545	ne	n_3	e
546	neh	n_3	eh
547	ng	ng_1	FNULL
548	nga	ng_2	a
549	ngai	ng_2	ai
550	ngau	ng_2	au
551	ngauh	ng_2	auh
552	nge	ng_3	e
553	ngeh	ng_3	eh
554	ngi	ng_4	i
555	ngia	ng_4	ia
556	ngiau	ng_4	iau
557	ngiauh	ng_4	iauh

558	ngiu	ng_4	iu
559	ngiuh	ng_4	iuh
560	ngo	ng_5	o
561	ngoeh	ng_5	oeh
562	ngou	ng_6	ou
563	ni	n_4	i
564	nia	n_4	ia
565	niau	n_4	iau
566	nih	n_4	ih
567	niou	n_4	iou
568	niu	n_4	iu
569	nng	n_8	ng
570	no	n_5	o
571	noa	n_5	oa
572	nou	n_6	ou
573	nui	n_7	ui
574	o	INULL_5	o
575	oa	INULL_5	oa
576	oah	INULL_5	oah
577	oai	INULL_5	oai
578	oaihN	INULL_5	oaihN
579	oaiN	INULL_5	oaiN
580	oaN	INULL_5	oaN
581	oan	INULL_5	oan
582	oat	INULL_5	oat
583	oe	INULL_5	oe
584	oeh	INULL_5	oeh
585	oh	INULL_5	oh

586	ok	INULL_5	ok
587	om	INULL_5	om
588	ong	INULL_5	ong
589	ou	INULL_6	ou
590	ouh	INULL_6	ouh
591	ouN	INULL_6	ouN
592	pa	p_2	a
593	pah	p_2	ah
594	pai	p_2	ai
595	paiN	p_2	aiN
596	pak	p_2	ak
597	paN	p_2	aN
598	pan	p_2	an
599	pang	p_2	ang
600	pat	p_2	at
601	pau	p_2	au
602	pauh	p_2	auh
603	pe	p_3	e
604	peh	p_3	eh
605	pek	p_3	ek
606	peN	p_3	eN
607	peng	p_3	eng
608	pha	ph_2	a
609	phah	ph_2	ah
610	phai	ph_2	ai
611	phaiN	ph_2	aiN
612	phak	ph_2	ak
613	phaN	ph_2	aN

614	phan	ph_2	an
615	phang	ph_2	ang
616	phau	ph_2	au
617	phauh	ph_2	auh
618	phe	ph_3	e
619	pheh	ph_3	eh
620	phehN	ph_3	ehN
621	phek	ph_3	ek
622	pheN	ph_3	eN
623	pheng	ph_3	eng
624	phi	ph_4	i
625	phiah	ph_4	iah
626	phiak	ph_4	iak
627	phiaN	ph_4	iaN
628	phian	ph_4	ian
629	phiang	ph_4	iang
630	phiat	ph_4	iat
631	phiau	ph_4	iau
632	phih	ph_4	ih
633	phin	ph_4	in
634	phiN	ph_4	iN
635	phio	ph_4	io
636	phit	ph_4	it
637	phiu	ph_4	iu
638	phng	ph_8	ng
639	phngh	ph_8	ngh
640	pho	ph_5	o
641	phoa	ph_5	oa

642	phoah	ph_5	oah
643	phoan	ph_5	oan
644	phoaN	ph_5	oaN
645	phoat	ph_5	oat
646	phoe	ph_5	oe
647	phoeh	ph_5	oeh
648	phoh	ph_5	oh
649	phok	ph_5	ok
650	phong	ph_5	ong
651	phou	ph_6	ou
652	phu	ph_7	u
653	phuh	ph_7	uh
654	phui	ph_7	ui
655	phun	ph_7	un
656	phut	ph_7	ut
657	pi	p_4	i
658	piah	p_4	iah
659	piak	p_4	iak
660	piaN	p_4	iaN
661	pian	p_4	ian
662	piang	p_4	iang
663	piat	p_4	iat
664	piau	p_4	iau
665	pih	p_4	ih
666	pin	p_4	in
667	piN	p_4	iN
668	pio	p_4	io
669	pit	p_4	it

670	piu	p_4	iu
671	png	p_8	ng
672	po	p_5	o
673	poa	p_5	oa
674	poah	p_5	oah
675	poaN	p_5	oaN
676	poan	p_5	oan
677	poat	p_5	oat
678	poe	p_5	oe
679	poeh	p_5	oeh
680	poh	p_5	oh
681	pok	p_5	ok
682	pong	p_5	ong
683	pou	p_6	ou
684	pu	p_7	u
685	puh	p_7	uh
686	pui	p_7	ui
687	puih	p_7	uih
688	puiN	p_7	uiN
689	pun	p_7	un
690	put	p_7	ut
691	sa	s_2	a
692	sah	s_2	ah
693	sahN	s_2	ahN
694	sai	s_2	ai
695	sak	s_2	ak
696	sam	s_2	am
697	saN	s_2	aN

698	san	s_2	an
699	sang	s_2	ang
700	sap	s_2	ap
701	sat	s_2	at
702	sau	s_2	au
703	se	s_3	e
704	seh	s_3	eh
705	sek	s_3	ek
706	seN	s_3	eN
707	seng	s_3	eng
708	si	s_4	i
709	sia	s_4	ia
710	siah	s_4	iah
711	siak	s_4	iak
712	siam	s_4	iam
713	sian	s_4	ian
714	siaN	s_4	iaN
715	siang	s_4	iang
716	siap	s_4	iap
717	siat	s_4	iat
718	siau	s_4	iau
719	sih	s_4	ih
720	sihN	s_4	ihN
721	sim	s_4	im
722	siN	s_4	iN
723	sin	s_4	in
724	sio	s_4	io
725	sioh	s_4	ioh

726	siok	s_4	iok
727	siong	s_4	iong
728	sip	s_4	ip
729	sit	s_4	it
730	siu	s_4	iu
731	siuN	s_4	iuN
732	sng	s_8	ng
733	sngh	s_8	ngh
734	so	s_5	o
735	soa	s_5	oa
736	soah	s_5	oah
737	soai	s_5	oai
738	soaiN	s_5	oaiN
739	soan	s_5	oan
740	soaN	s_5	oaN
741	soat	s_5	oat
742	soe	s_5	oe
743	soeh	s_5	oeh
744	soh	s_5	oh
745	sok	s_5	ok
746	som	s_5	om
747	song	s_5	ong
748	sou	s_6	ou
749	su	s_7	u
750	suh	s_7	uh
751	sui	s_7	ui
752	suiN	s_7	uiN
753	sun	s_7	un

754	sut	s_7	ut
755	ta	t_2	a
756	tah	t_2	ah
757	tai	t_2	ai
758	taiN	t_2	aiN
759	tak	t_2	ak
760	tam	t_2	am
761	tan	t_2	an
762	taN	t_2	aN
763	tang	t_2	ang
764	tap	t_2	ap
765	tat	t_2	at
766	tau	t_2	au
767	tauh	t_2	auh
768	te	t_3	e
769	teh	t_3	eh
770	tek	t_3	ek
771	teN	t_3	eN
772	teng	t_3	eng
773	tha	th_2	a
774	thah	th_2	ah
775	thai	th_2	ai
776	thaiN	th_2	aiN
777	thak	th_2	ak
778	tham	th_2	am
779	thaN	th_2	aN
780	than	th_2	an
781	thang	th_2	ang

782	thap	th_2	ap
783	that	th_2	at
784	thau	th_2	au
785	the	th_3	e
786	theh	th_3	eh
787	thek	th_3	ek
788	theN	th_3	eN
789	theng	th_3	eng
790	thi	th_4	i
791	thiah	th_4	iah
792	thiam	th_4	iam
793	thiaN	th_4	iaN
794	thian	th_4	ian
795	thiap	th_4	iap
796	thiat	th_4	iat
797	thiau	th_4	iau
798	thih	th_4	ih
799	thim	th_4	im
800	thin	th_4	in
801	thiN	th_4	iN
802	thio	th_4	io
803	thiok	th_4	iok
804	thiong	th_4	iong
805	thit	th_4	it
806	thiu	th_4	iu
807	thng	th_8	ng
808	tho	th_5	o
809	thoa	th_5	oa

810	thoah	th_5	oah
811	thoaN	th_5	oaN
812	thoan	th_5	oan
813	thoat	th_5	oat
814	thoe	th_5	oe
815	thoeh	th_5	oeh
816	thoh	th_5	oh
817	thok	th_5	ok
818	thong	th_5	ong
819	thou	th_6	ou
820	thu	th_7	u
821	thuh	th_7	uh
822	thui	th_7	ui
823	thuiN	th_7	uiN
824	thun	th_7	un
825	thut	th_7	ut
826	ti	t_4	i
827	tia	t_4	ia
828	tiah	t_4	iah
829	tiak	t_4	iak
830	tiam	t_4	iam
831	tiaN	t_4	iaN
832	tian	t_4	ian
833	tiang	t_4	iang
834	tiap	t_4	iap
835	tiat	t_4	iat
836	tiau	t_4	iau
837	tih	t_4	ih

838	tihN	t_4	ihN
839	tim	t_4	im
840	tiN	t_4	iN
841	tin	t_4	in
842	tio	t_4	io
843	tioh	t_4	ioh
844	tiok	t_4	iok
845	tiong	t_4	iong
846	tip	t_4	ip
847	tit	t_4	it
848	tiu	t_4	iu
849	tiuh	t_4	iuh
850	tiuN	t_4	iuN
851	tng	t_8	ng
852	to	t_5	o
853	toa	t_5	oa
854	toah	t_5	oah
855	toaN	t_5	oaN
856	toan	t_5	oan
857	toat	t_5	oat
858	toe	t_5	oe
859	toh	t_5	oh
860	tok	t_5	ok
861	tom	t_5	om
862	tong	t_5	ong
863	tou	t_6	ou
864	touh	t_6	ouh
865	tu	t_7	u

866	tuh	t_7	uh
867	tui	t_7	ui
868	tuiN	t_7	uiN
869	tun	t_7	un
870	tut	t_7	ut
871	u	INULL_7	u
872	uh	INULL_7	uh
873	ui	INULL_7	ui
874	uih	INULL_7	uih
875	uiN	INULL_7	uiN
876	un	INULL_7	un
877	ut	INULL_7	ut

附錄二

FA40 嬒	FA41 㑳	FA42 嬡	FA43 迌	FA44 𨑨	FA45 刣	FA46 啉	FA47 躼
FA48 哰	FA49 嗵	FA4A 頼	FA4B 哊	FA4C 婧	FA4D 穲	FA4E 揞	FA4F 嫺
FA50 痟	FA51 趙	FA52 炁	FA53 偲	FA54 偲	FA55 捆	FA56 撩	FA57 瘖
FA58 酏	FA59 褫	FA5A 趣	FA5B 擇	FA5C 勢	FA5D 禮	FA5E 擨	FA5F 揥
FA60 揬	FA61 樸	FA62 鯀	FA63 佮	FA64 搝	FA65 喓	FA66 捌	FA67 揤
FA68 躁	FA69 踮	FA6A 徛	FA6B 瘜	FA6C 爁	FA6D 敊	FA6E 組	FA6F 咗
FA70 隶	FA71 坮	FA72 瘨	FA73	FA74	FA75 嗌	FA76 擳	FA77 轂
FA78 唔	FA79 覨	FA7A 胶	FA7B	FA7C 蹄	FA7D 踂	FA7E	↵
	FAA1 襉	FAA2	FAA3	FAA4	FAA5 揁	FAA6	FAA7 瞷
FAA8	FAA9	FAAA	FAAB	FAAC 簕	FAAD 捒	FAAE 鈎	FAAF
FAB0 喰	FAB1 徭	FAB2	FAB3 甥	FAB4	FAB5 扲	FAB6	FAB7 僐
FAB8 威	FAB9 樣	FABA	FABB	FABC	FABD	FABE	FABF
FB50 堃	FB51	FB52	FB53	FB54	FB55	FB56	FB57
FCE8	FCE9	FCEA	FCEB	FCEC	FCED 增	FCEE	FCEF
FCF0 媸	FCF1 剾	FCF2 侹	FCF3 旨	FCF4 躿	FCF5 噭	FCF6 滘	FCF7 埪
FCF8 膣	FCF9 启	FCFA 糁	FCFB 婌	FCFC 熣	FCFD 嗦	FCFE 㴁	↵
FD40 幰	FD41 熻	FD42 瘞	FD43 痾	FD44 蹤	FD45 躪	FD46 枅	FD47 膈
FD48 犷	FD49 甂	FD4A 舤	FD4B 烺	FD4C 胖	FD4D 乜	FD4E 濬	FD4F 撳
FD50 搭	FD51 攖	FD52 掿	FD53 擭	FD54 攜	FD55 揩	FD56 撓	FD57 搠
FD58	FD59 揵	FD5A	FD5B 捶	FD5C 捲	FD5D 揠	FD5E 个	FD5F 旒
FD60	FD61 婔	FD62 脝	FD63 挔	FD64 黷	FD65 瘃	FD66 龕	FD67 虄
FD68 凭	FD69 不	FD6A 嚨	FD6B 餾	FD6C 刣	FD6D 矚	FD6E 犽	FD6F 偲
FD70 綮	FD71	FD72 呷	FD73 壆	FD74	FD75 誐	FD76	FD77 諂
FD78 跙	FD79 糟	FD7A 黏	FD7B	FD7C 絚	FD7D 迋	FD7E 噍	↵
	FDA1 眉	FDA2 朏	FDA3 唛	FDA4 撥	FDA5 緆	FDA6	FDA7 囓
FDA8	FDA9	FDAA	FDAB	FDAC	FDAD	FDAE	FDAF
FDB0 愺	FDB1 滛	FDB2 餺	FDB3 嫐	FDB4 冋	FDB5 跤	FDB6 棠	FDB7 籸
FDB8 抄	FDB9 揸	FDBA 搘	FDBB 摟	FDBC 揙	FDBD 蹽	FDBE 劊	FDBF 劇
FDC0 瘃	FDC1 晡	FDC2 啾	FDC3 穑	FDC4	FDC5	FDC6	FDC7