以知識表徵方法建構台語聲調群剖析器[[1]](#footnote-1)

A Knowledge Representation Method to Implement   
A Taiwanese Tone Group Parser

張佑竹[[2]](#footnote-2)\*

Yu-Chu Chang

摘要

聲調群剖析器是台閩語語音輸出系統的主要元件之一。本文提出聲調管轄假說，主張先將句內語詞定調，亦能決定台閩語聲調群分界的觀點，並以聲調群剖析器實作加以驗證。除了敘述如何應用預設調型、預設詞類和模式三種標記符號，將語言知識和經驗轉換為知識庫，並說明經由推論引擎與知識庫的連結，完成語詞定調的運作過程。目前內部測試平均變調正確率為98.5%。外部測試平均變調正確率為94%。本研究的實驗數據也顯示一個重要的線索：符號系統標記比規則推論對變調正確性有相對較高的貢獻率。

關鍵詞：台灣話，變調，聲調群剖析器，知識表徵，模擬

Abstract

A tone group parser could be one of the most important components of the Taiwanese text-to-speech system. In this paper, we offered the hypothesis of tonal government to emphasis the idea that if the allotone selection can be made for each word in a sentence then the tone groups will be separated within the sentence and supported our viewpoint with the implementation of a Taiwanese tone group parser. In addition to the description of using the symbol system to convert language expertise and heuristic knowledge into a knowledge base to cope with a frame-based corpus and a tone sandhi processor, the procedure of connecting the inference engine and the knowledge base to make allotone selection was also discussed. In the current version of the tone group parser, the average accuracy of inside test is 98.5%. The average accuracy of outside test is 94%. The experiment data of the study also reveals an important clue: the marking of the symbol system makes a higher contribution rate to the tone sandhi accuracy than the rule inference.

Keywords: Taiwanese, Tone Sandhi, Tone Group Parser, Knowledge Representation, Simulation

# 緒論 (English)

聲調群(Tone groups)是台閩語[[3]](#footnote-3)的基本韻律結構。聲調群剖析器也是台閩語語音輸出系統的主要元件之一(Liim, 2004; 田村志津枝, 2010)。本文首先從音韻和語言結構的觀點探索台閩語的特質，將台閩語如何藉著變調(Tone sandhi)形成獨特聲調群結構的衍生過程和當代音韻-句法界面的相關研究相互印證，提出聲調管轄假說(Tonal-government hypothesis)來主張先將句內語詞定調，亦能決定聲調群分界的觀點。隨後說明以台閩語變調習得(Tone sandhi acquisition)理論為基礎，將知識表徵(Knowledge representation)技術和語詞屬性分析加以整合，製作聲調群剖析器的方法。

# 文獻回顧 (English)

## 2.1 從音韻和語言結構觀點看台閩語特質 (English )

台閩語為聲調語言。變調是指語詞聲調受到相鄰語詞影響而改變聲調調值的現象，常見於中國各地的語言。比較特別的是台閩語語詞具有普遍的變調現象。每個台閩語語詞皆有本調[[4]](#footnote-4)(Lexical tone)和變調(Sandhi tone)兩種調型(Tone form)。語詞或聲調群的最後音節讀本調，其餘音節讀變調 (Chiu, 1931; 王育德, 1955)。因此，若且唯若一個或一組語詞中僅有最後一個音節讀本調，則此一語詞或詞組即為聲調群。換句話說，台閩語語句就是聲調群的集合。聲調群不僅是組成台閩語語句的句法單元(Syntax unit)，同時也是完整的語義單位(Semantic unit)和韻律結構(Prosodic structure)。台閩語很可能是唯一在語句內以變調方式建立聲調群結構的自然語言(Chang, 2009)。

### 2.1.1 台閩語變調與聲調群的形成 (Taiwanese Tone Sandhi and the Formation of Tone Group)

台閩語的聲調、語義和句法之間有密切的關係。聽話者可以從相同的語句中以聲調來辨別不同的詞類(POS)和語義。例如，本調相同的「ke55(雞/加)」在例句(1)和(2)裡，有不同的調型、詞類和語義。

(1) Tsit31 tsiah31（變調）ke55（本調，名詞）tsit54 kong55-kin55.（這隻雞重一公斤。）

(2) Tsit31 tsiah31（本調）ke55（變調，動詞）tsit54 kong55-kin55.（這隻多重一公斤。）

這個例句讓我們注意到人腦雖然可以就台閩語同音異形漢字「雞/加」和語境來分析語義和句法結構，但是對電腦而言，(1)和(2)的羅馬拼音完全相同，只有自主語義(Autonomous semantic mapping)確定以後，才能決定「ke55」的調型或進行句法分析。這部分屬於高階人工智能(Strong AI)的範疇，也是對話系統必須面對的困境。

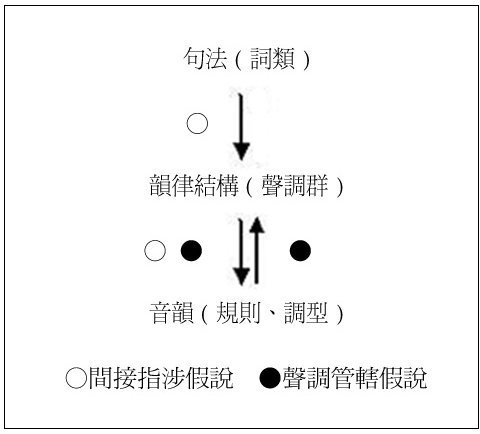


圖1. 台閩語句法、韻律結構和音韻間的關係及相關假說的適用範域  
[Figure 1. The relationship among syntax, prosodic structure and phonology in Taiwanese with the related hypotheses]

***表1. 台閩語語詞預設調號及處理方式  
[Table 1. The list of default mark of tone form]***

|  |  |  |  |
| --- | --- | --- | --- |
| 預設調型記號 | 預設調型 | 適用語詞 | 處理方式 |
| 0 | 固定讀本調 | 只讀本調的語詞 | 不需推論 |
| 1 | 預設讀變調 | 單音節語詞 | 以規則推論 |
| 2 | 預設讀本調 | 預設讀本調的語詞如詞組，輕聲詞，外國語 | 以規則推論 |
| 3 | 固定讀變調 | 只讀變調的語詞 | 不需推論 |
| # | 本調或變調 | 可能讀本調或變調的語詞 | 以規則推論 |
| & | 固定讀本調 | 聲調群或聲調群集合 | 不需推論 |

參考文獻(References)

Pan, H. H. (2003). Prosodic hierarchy and nasalization in Taiwanese. In *Proceedings of the 15th ICPhS*, 575-578.

Selkirk, E. O. (1986). *Phonology and syntax: the relationship between sound and structure*. Cambridge, MA: MIT press.

Tsay, J. (1999). Bootstrapping into Taiwanese tone sandhi. In *Chinese Languages and Linguistics V, Symposium Series of the Institute of History and Philology*, *2*(5), 311-333. Taipei, Taiwan: Academia Sinica.

Tsay, J., Myers, J., & Chen, X. J. (2000). Tone sandhi as evidence for segmentation in Taiwanese. In *Proceedings of the 30th Child Language Research Forum*. 211-218.

王育德（1955）。台灣語の聲調。中國語學，41，3-11。[Ong, I.T. (1955). Taiwanese Tones. Journal of Chuugoku Gogaku, 41, 609-617.]

田村志津枝（2010）。初めて台湾語をパソコンに喋らせた男―母語を蘇らせる物語。東京：現代書館。[Tamura, S.(2010).Hajimete Taiwango o pasokon ni shaberaseta otoko: bogo o yomigaeraseru monogatari.Tokyo, Japan: Gendai Shokan.]

1. 本論文承蒙中正大學語言學研究所麥傑教授提供諸多建議，謹此致謝。 [↑](#footnote-ref-1)
2. \* 國立中正大學語言學研究所  
   Institute of Linguistics, National Chung Cheng University  
   E-mail: xxxx@xxxxx [↑](#footnote-ref-2)
3. 台灣話源自中國閩南方音。公元2006年教育部公告台灣閩南語羅馬字拼音方案，台灣話得以文 字公開傳承。本文例句，採用教育部台灣閩南語（以下簡稱台閩語）拼音加註聲調值。引述論文中的Taiwanese或台灣語也併譯為台閩語。 [↑](#footnote-ref-3)
4. 本調Lexical tone亦稱juncture tone，變調sandhi tone亦稱context tone。 [↑](#footnote-ref-4)