

主要漢字形聲字發音規則探勘與視覺化

Primary Chinese Semantic-Phonetic Compounds Pronunciation Rules

Mining and Visualization

徐千惠 Chien-Hui Hsu

國立中央大學資訊工程系

Department of Computer Science and Information Engineering

National Central University

shu252000@gmail.com

蔡孟峰 Meng-Feng Tsai

國立中央大學資訊工程系

Department of Computer Science and Information Engineering

National Central University

mftsai@csie.ncu.edu.tw

張嘉惠 Chia-Hui Chang

國立中央大學資訊工程系

Department of Computer Science and Information Engineering

National Central University

chia@csie.ncu.edu.tw

廖湘美 Hsiang-Mei Liao、李淑萍 Shu-Ping Li,

國立中央大學中國文學系

Department of Chinese Literature

National Central University

anne54555@gmail.com, leesp.susan@gmail.com

吳嫻 Denise H. Wu

國立中央大學認知神經科學研究所

College of Science Institute of Cognitive Neuroscience

National Central University

denisewu@cc.ncu.edu.tw

摘要

近年華語教學的需求量與重要性日漸增加，為幫助漢語學習者構築現代漢字、增進學習效率，採用「部件教學法」，由部件當中找出漢字表音和表意的線索，故以形聲字與其聲符為研究對象。形聲字在現代漢語通用字中佔八成，大多是由一個表意的形符加上一個表音的聲符組成。本研究強調聲符表音的線索，以關聯式規則探勘出形聲字發音規則。並進一步地找出影響形聲字發音的關鍵因素，輔以漢語音韻學的知識，建立漢字發音的階層架構，進行多層次形聲字發音規則探勘，藉此幫助漢語學習者與教學研究歸納形聲字發音的脈絡。最後用視覺化的方式呈現這些規則，並設計簡單、好記的系統輔助漢字識字教學與漢字研究。

Abstract

The demand and the importance of Chinese teaching have increased continuously. In order to assist the Chinese learners in composing Chinese characters and increase their learning efficiency, Chinese components teaching method is adopted. The learners can find the clues to both the pronunciations and the meanings of Chinese characters from Chinese components, and semantic-phonetic compounds and their phonetic components are exactly proper to be the object. There are 80.5% semantic-phonetic compounds in the 7000 common Chinese characters, and most of them are formed with one semantic component and one phonetic component. For the purpose of emphasizing the clues to the pronunciations of Chinese characters, multiple-level association rule mining was applied to discover the hierarchical pronunciation rules of semantic-phonetic compounds. This approach found the key factors which have the strong connection with the pronunciations of semantic-phonetic compounds. With the knowledge of Chinese linguistics, we constructed the hierarchical Chinese pronunciation structure. The hierarchical pronunciation rules are the overview of the pronunciations of semantic-phonetic compounds and aid both Chinese learning and Chinese researches. Therefore, they can learn the pronunciations of Chinese characters not only in the general aspect but the specific aspect. These rules were represented in visualization and the simple and memorable system was designed to assist both the Chinese literacy teaching and Chinese researches.

關鍵詞：漢語識字教學，形聲字，聲符，多層次關聯式規則探勘，關聯式規則視覺化

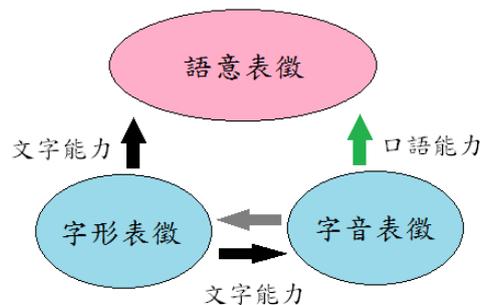
Keywords: Chinese Teaching, Semantic-Phonetic Compounds, Phonetic Component, Multiple-level Association Rule Mining, Association Rule Mining Visualization.

一、緒論

漢字為一種歷史悠久的形系文字，現今全球有五分之一的人口使用漢語做為母語，再加上華人社會經濟、文化影響力的擴大與市場開發的需要，愈來愈多人將漢語做為第二外語來學習，大約有 100 個國家超過 2,500 所大學教授漢語課程[1]，成為最多人學習的東方語言，且學習人數益形增加。進一步來看，現今海外華人人數約為五千萬人；而在台灣的大陸與外籍配偶從 2002 年的二十二萬人次，已增加到四十四萬人次，其中外籍配偶約占十四萬六千人次，這些華語學習者雖接觸到漢語環境，但識字對於他們來說仍感到相當困難。

外國的漢語學習者遇到最大的難處在於漢字的學習。由於漢字的「形」不直接表音的特性，使得漢字的發音教學深具難度。漢字教學應當有別於占絕大優勢的拼音文字教學法，若是忽略了漢字的組字特性，學生只好靠著經驗想辦法消化所學的漢字，喪失了漢字中表音與表意的線索，造成漢語學習者在學習過程中更加困難重重。同時，教師要教會學生會聽、會說、會讀、會寫漢字，於是漢語學習者要學會一個漢字耗時且耗力[2]。

語言學習的過程如圖一所示，學習一種語言可分成學習其字音、字形與字義三部分，這三部分在學習者的大腦形成的印象就稱為表徵，口語能力屬於字音表徵對應到語意表徵的過程，文字能力則屬於字形表徵對應到字音表徵和語意表徵的過程，甚至是字音表徵對應到字形表徵的過程。對於海外華人和外籍配偶來說，在日常生活中已接觸到漢語對話，自然而然有基本的口語能力，而漢字的文字能力才是他們需要加強的部份。另外對於華人學童與閱讀障礙者，加強他們的文字能力，更可以幫助他們將所學運用在漢語口語和閱讀當中。不過，漢字字形不直接表音的特性，使得字形對應到字音往往需要以拼音來輔助，相較於拼音文字系統來說，漢字系統多了一步學習的過程，為學習者築起一道鴻溝。若能設計合適的漢字識字教學，指引漢語學習者找回漢字表音和表意的線索，將有助於他們跨越漢字學習的障礙，增加其學習效率的同時亦能窺探漢字的奧秘。



圖一、語言學習的過程

反觀屬於拼音文字系統的英語，若是英語學習者同樣具備基本的英語口語能力，欲進一步加強其文字能力就顯得輕鬆許多。這正是因為學會英語的拼音系統後，學習者就能自然而然地邊念單字邊把它拼寫(spell)出來。

以歷史的演進來看，漢字可以用傳統的「六書理論」，將漢字分類成：象形、形聲、會意、指事、轉注、假借六種造字方法[3]。但是語言是有生命的東西，現代漢字的形體早有變化，故要以「六書理論」來分類現代漢字並不容易，亦不適合教導給漢語學習者。於是，學者提出漢字部件教學的概念。所謂的漢字部件，即「現代漢字字形中，具有獨立組字能力的構字單位，它大於或等於基本筆畫，小於或等於整個漢字」[4]，如「好」字，就是由「女」+「子」兩個部件所組成。以漢字部件幫助漢語學習者構築現代漢字、學習到部件延伸的漢字，是為基礎且有效率的教學方法，稱為「部件教學法」。相關研究也發現以部件為主之教學可以提升新移民女性的漢字學習成效[5]，尤其成人學漢字，著重在認知轉換，並善用其理解和歸納的能力[6]。由此方法找出漢字表音和表意的線索，形聲字是最能被直接觀察到此種特徵的漢字結構。在 7000 個現代漢語通用字當中[7]，形聲字佔了 80.5%，共 5631 個形聲字；又，在前 3000 個現代漢字常用字當中，形聲字佔了 57.4%，共 1721 個形聲字。大部分的形聲字是兩個部件組合在一起，即一個表意的形符加一個表音的聲符。若能強調由聲符表音的線索，將能給予中文研究與教學大幅度的進展，提供中文研究一套有系統性的研究成果，並在漢語教學的過程中，讓

學習者由學習到的生字，增加文字解析歸納的能力並延伸學習到相同線索的漢字，減少記憶的負擔，進而提升漢字學習效率。

爲了強調聲符表音的線索並找出其重要規則，首先採用中央資工所與中文所的四位研究生與三位教授合作所建立的形聲字資料庫。此形聲字資料庫是運用中研院文獻處理實驗室所建立的「漢字構形資料庫」[8]，加上自創的形聲字源標記系統，由中文所師生人工標記在這漢字構形資料庫當中，含有注音標示的 14598 個漢字是否爲形聲字、其聲符爲何等資訊，並耗時兩年完成。再對形聲字資料庫進行影響形聲字發音的因素分析，在所有資料的屬性中，找出最能影響形聲字發音的屬性，並以此作爲規則探勘的項目。本研究採用注音符號與漢語拼音，並輔以漢語音韻學的知識，將漢字發音分成三個層次來看。進而以多層次探勘形聲字發音規則，以此幫助漢語學習者與教學研究歸納形聲字發音的情形、了解漢字發音的脈絡，並找出「主要的形聲字發音規則」。我們發現，雖然大部分的形聲字發音與其聲符發音相同，但主要的形聲字發音規則也揭露不少例外的情形，我們也將深入分析。進一步地，以視覺化的方式呈現規則，更能幫助學習者一目了然且容易記憶發音規則，亦可協助教學研究者加以分析比較規則的內容。並設計注音符號版本與漢語拼音版本的互動式網站系統，再輔以範例字，擴展漢語學習者的識字量，同時作爲漢字研究上的佐證。

接下來的論文架構爲：第二章介紹與此論文相關的方法與研究內容，第三章敘述如何在影響形聲字發音的因素分析中，找出形聲字資料庫中最具影響力的屬性(一)，並加上漢語音韻學的知識，建立漢字發音階層架構(二)，以此找出多層次發音關聯規則(三)，設計符合規則的視覺化方法(四)。第四章呈現主要探勘的成果與網站內容，第五章爲結論與未來方向。

二、相關研究

(一) 漢語教學研究

由於漢語學習市場的擴增，各式各樣的漢語教學法紛紛出爐，但如何有效學習漢語、由淺入深地掌握學習漢語的要領，減少外國學習者學習漢語的挫折感，是漢語教學的關鍵。林季苗認爲，漢字的形音義間的關係和拼音文字存在著極大的差異，故漢字教學應當有別於占絕大優勢的拼音文字教學法[2]。林季苗提出四項漢語教學原則與在法國的漢語教學經驗，四項教學原則之一爲「字本位」，強調每個字本身的意義、構型與發音。原則二爲「語文分步」，說明漢語的口語及文字教學應當適當地分開進行，保持學生行文的順暢。原則三爲「集中識字」，強調有系統、有目的地、循序漸進將漢字由淺入深教授給學生。原則四是「區別主動書寫字及被動認讀字」，讓學生可專心將精神與時間著重在基礎漢字或其它漢語能力上。

中研院李佳穎博士針對漢字識字教學，採用聲符部件所延伸的漢字做爲集中識字的方法，並以字本位教授學童這些漢字的組字特性，輔助中文學習者識字[9]。另外，李博士以大腦認知的角度，測量學童在辨識不同特性的形聲字的識字速度，並發現影響形聲字識字速度的特性大致上分爲：頻率、一致性和規則性。所謂頻率是指一個字在日常生活中出現的次數，一致性爲同個聲符的所有形聲字之間發音相似的程度，規則性是指形聲字與其聲符兩者之間發音相似的程度。對於頻率高的漢字，不管是否易學，學童識字

表 一、聲母分類表(左欄)與韻母分類表(右欄)

發音方法 發音部位	塞音	塞擦音	擦音	邊音	鼻音
雙脣	ㄅ ㄆ			ㄇ	
脣齒			ㄈ		
舌尖中	ㄉ ㄊ			ㄋ	ㄌ
舌尖前		ㄗ ㄘ	ㄙ		
舌尖後		ㄑ ㄒ	ㄖ ㄛ		
舌面		ㄐ ㄑ	ㄔ		
舌根	ㄍ ㄎ		ㄗ		

韻別 嘴型	單韻	複韻	聲隨韻	捲舌韻
開口	ㄚ ㄛ ㄜ ㄝ ㄞ ㄟ	ㄝ ㄜ ㄝ ㄞ ㄟ	ㄛ ㄜ ㄝ ㄞ ㄟ	ㄝ ㄞ ㄟ
齊齒	一	結合韻母		
合口	ㄨ	ㄨ ㄚ, ㄨ ㄛ, ㄨ ㄜ, ㄨ ㄝ, ㄨ ㄞ, ㄨ ㄟ, ㄨ ㄛ, ㄨ ㄜ, ㄨ ㄝ, ㄨ ㄞ, ㄨ ㄟ		
撮口	ㄨ	ㄨ ㄛ, ㄨ ㄜ, ㄨ ㄝ, ㄨ ㄞ, ㄨ ㄟ		

(三) 關聯式規則探勘與視覺化

資料探勘被定義為從資料進行知識發掘(Knowledge Discovery from Data)的過程中，以智慧的方式擷取資料樣式。關聯式規則探勘出現於 1990 年代[12]，原是用於購物籃分析，在顧客交易資料庫中，觀察購買項目間隱含的關係，了解顧客的消費習慣，例如：

牛奶⇒麵包

此例代表買牛奶的顧客也傾向在購買期間內選購麵包。為測量這些隱含規則，提出普遍性(support)和正確性(confidence)的測量標準，表示如下：

$$\text{普遍性}(A \Rightarrow B) = P(A \cup B) = \frac{\text{資料筆數}(A \cup B)}{\text{全部資料筆數}} \tag{1}$$

$$\text{正確性}(A \Rightarrow B) = P(B | A) = \frac{\text{普遍性}(A \cup B)}{\text{普遍性}(A)}$$

D 代表交易資料庫，A、B 各代表一個個體或群體的項目集(item-set)。A⇒B 的普遍性代表在資料庫 D 中同時出現項目 A 與 B 的比例，以機率 P(A∪B) 表示。A⇒B 的正確性代表在資料庫 D 中，如果已經出現項目 A 時，項目 B 也同時出現的比例，以條件機率 P(B|A) 表示。在進行關聯規則探勘時，可先設定最小普遍性(minimum support)與最低正確性(minimum confidence)，做為強關聯式規則的門檻。

關聯式規則視覺化可幫助決策者加以分析，其相關研究有：散播平面圖(scatter plot)、以圖解為基礎的視覺化(graph-based visualization)、平行座標圖(parallel coordinates plots)、雙層圖(double decker plot)、以矩陣為基礎的視覺化(matrix-based visualization)[13]並加以分群[14]等。

三、形聲字重要發音規則探勘與視覺化

本研究幫助中文學習者加強漢字識字能力，設計一套合適的漢字識字輔助教學系統，幫助他們由漢字的組字特性中，加深其聲符表音概念的形成；亦提供漢字教學研究加以運用。本研究分成四個階段，首先是取得形聲字的相關資料並進行影響形聲字發音的因素

分析，再來建立漢字發音的階層式架構，找出主要的形聲字發音規則，最後設計視覺化方法、以教學網站的方式呈現。

(一) 影響形聲字發音的因素分析

首先，本研究的形聲字資料是沿用國立中央大學中文所與資工所師生合作所建立的形聲字資料庫。他們應用中研院文獻處理實驗室建立的「漢字構形資料庫」，建立形聲字標記系統，再由中文所四位研究生與三位教授人工標記形聲字與其聲符，最後耗時兩年多將所有含注音標示的 14598 個漢字標記完成。此形聲字資料庫共記錄了 9292 個形聲字、1431 個聲符。

在探勘形聲字發音規則之前，首先分析形聲字的發音特性，我們發現：有 55.5% 的形聲字的發音與其聲符的聲母、韻母皆相同，在另外 44.5% 的情況下，有哪些屬性最可以協助我們辨別形聲字的發音？本研究採用 Mutual Information (互信息)[15] 來計算每個屬性對於形聲字聲母、韻母的影響程度，其公式內容為資訊熵減去條件熵，如 (2) 式所示。

$$I(X;Y) = H(X) - H(X/Y) = - \sum_{x \in X} \Pr(x) \log \Pr(x) + \sum_{y \in Y} \sum_{x \in X} \Pr(y,x) \log \Pr(x/y) \quad (2)$$

當中可能影響形聲字發音的屬性與屬性值列表於表二中，其中聲符的韻母=「ㄇ」代表發音為空韻。舉例屬性影響形聲字發音：欲計算「連接方式」影響聲符韻母=一ㄇ的形聲字發音的 Mutual Information 值，先列出符合條件的字數如表三，表中的 285 字表示當形聲字的聲符韻母=一ㄇ且形聲字韻母=一ㄇ時，符合上下連接的形聲字有 285 個字。之後將此表的資訊套入(2)式：

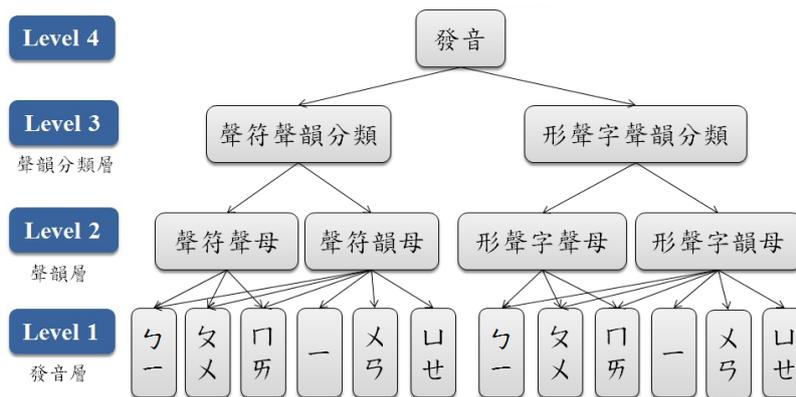
$$I(X;Y) = -\left(\frac{361}{458} \log \left(\frac{361}{458}\right) + \frac{97}{458} \log \left(\frac{97}{458}\right)\right) - \left(\frac{285}{458} \log \left(\frac{285}{458} / \frac{357}{458}\right) + \left(\frac{52}{458} \log \left(\frac{52}{458} / \frac{65}{458}\right) + \left(\frac{24}{458} \log \left(\frac{24}{458} / \frac{36}{458}\right) + \left(\frac{72}{458} \log \left(\frac{72}{458} / \frac{357}{458}\right) + \left(\frac{13}{458} \log \left(\frac{13}{458} / \frac{65}{458}\right) + \left(\frac{12}{458} \log \left(\frac{12}{458} / \frac{36}{458}\right)\right) \approx$$

0.008。式子中 $\frac{361}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其韻母=一ㄇ的比例為 $\frac{361}{458}$ ； $\frac{97}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其韻母=其它的比例為 $\frac{97}{458}$ ； $\frac{285}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其韻母=一ㄇ且連接方式=上下連接的比例為 $\frac{285}{458}$ ； $\frac{357}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其連接方式=上下連接的比例為 $\frac{357}{458}$ ； $\frac{52}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其韻母=一ㄇ且連接方式=左右連接的比例為 $\frac{52}{458}$ ； $\frac{65}{458}$ 代表在所有符合聲符韻母=一ㄇ的形聲字當中，其連接方式=左右連接的比例為 $\frac{65}{458}$ ，其它數值對照表三依此類推...，計算所得的 Mutual Information

值為 0.008，代表屬性「連接方式」與聲符發音關聯性弱。若是 Mutual Information 值越大代表此屬性與聲符發音關聯性越強，而所有屬性影響形聲字聲母、韻母是否與其聲符聲母、韻母相同的 Mutual Information 值於圖二表示。圖二橫軸是形聲字以其聲符的聲母與韻母為分類，縱軸表示每個屬性對於不同分類的 Mutual Information 值，其中屬性依序以所代表的顏色顯示在長條圖中。由圖中可看出：影響形聲字的發音是否與聲符發音相同的因素在於聲符的聲母和韻母(藍色與靛色的部分最長)，以此作為形聲字發音規則的探勘項目，提供漢語學習者重要關鍵的形聲字發音規則。

(二) 漢字發音階層架構

漢字的發音教學往往需要拼音輔助，此拼音可為注音符號、漢語拼音或國際音標，例如注音符號「ㄉ」，在漢語拼音中表示成「d」，在國際音標中以「d」代表，本研究採用注音符號與漢語拼音兩種表示漢字的發音。在漢語音韻學當中，將漢語的發音分成聲母、韻母和聲調三部分，進一步地，不同聲母又可依照發音方法、發音部位兩種方法來分類，而不同韻母可依照韻別、嘴型兩種方法進行分類。故本研究採取在不同層次上表達漢字的發音，參照漢語聲韻學發音分類，定義形聲字與其聲符發音的階層式架構，如圖三所示。圖三顯示階層式架構的根結點為第四層，代表所有漢字的發音。圖中右邊分支代表形聲字的發音，左邊分支代表其聲符的發音。在第三層是發音的聲母與韻母的分類，其中有「發音方法」、「發音部位」、「韻別」和「嘴型」四種分類方法，稱為聲韻分類層；在第二層記錄發音的聲母、韻母，為聲韻層；在第一層則是完整的發音(忽略聲調)，稱作發音層，以此表示本研究在不同層次上探勘形聲字發音規則的概念。表四列出由第一層到第三層的發音階層結構例子。使用階層式架構的目的，是為幫助使用者由不同發音細微度 (granularities) 學習形聲字發音規則，其發音細微度意指發音單位大小，以此幫助漢語學習者歸納形聲字發音的情形，帶領他們從宏觀或是細微的角度學習漢字發音的脈絡。



圖三、漢字發音階層架構圖

表四、發音階層結構表示法

Level 1 發音層	Level 2 聲韻層		Level 3 聲韻分類層			
形聲字	聲母	韻母	發音方法	發音部位	韻別	嘴型
ㄉ一ㄠ/diao(刁)	ㄉ/d	一ㄠ/yao	塞音	舌尖中	結合韻母	齊齒
尸ㄛ/she(什)	尸/shi	ㄛ/e	擦音	舌尖後	單韻	開口
聲符	聲母	韻母	發音方法	發音部位	韻別	嘴型
ㄉ一ㄠ/diao(刁)	ㄉ/d	一ㄠ/yao	塞音	舌尖中	結合韻母	齊齒
尸/shi(十)	尸/shi		擦音	舌尖後	單韻	開口

母「發音方法」發音關聯規則的視覺化，可看出形聲字的聲符「發音方法」大多數和形聲字的「發音方法」相同。透過規則的視覺化，學習者一眼就可看出規則涵蓋的範圍與其重要程度，漢字研究者亦可加以分析比較，不但能深入了解發音規則，亦可概觀整體聲符發音影響形聲字發音的情形，完整呈現形聲字發音的脈絡、易於記憶發音規則。

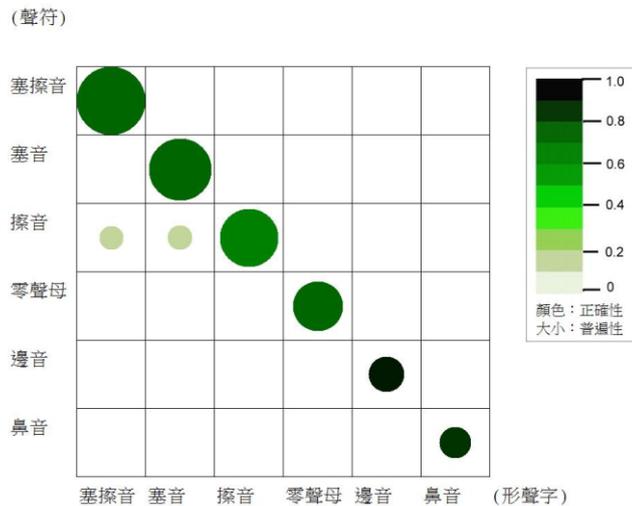


圖 四、規則視覺化-Level3 「發音方法」

四、主要貢獻成果

本研究在 9292 個形聲字資料庫中，將最小正確性設為 0.1，最小普遍性設為 0.001，探勘出強關聯規則，並且揭露形聲字發音與其聲符發音不同的轉音規則。篩選規則的條件為：在同個發音階層上的「形聲字的聲符發音 → 形聲字發音」，所得的規則稱為「主要的形聲字發音規則」。表六列出在第一層到第三層發音結構上探勘出的主要形聲字發音規則數量，並將探勘產生的部分強關聯規則於表七、表八中呈現，表八中 ID=1 的聲母=空，其意思為無聲母。舉列表七中 ID=1 如下：

聲符聲母=舌面 → 字聲母=舌面，普遍性=15%，正確性=75%

其含義為：在 9292 個形聲字中，符合形聲字的聲符聲母的發音部位為舌面的條件下，有 75% 形聲字聲母的發音部位也是舌面，共佔 15%。我們可看出：表八的發音關聯規則，其前八項顯示形聲字的發音部位與其聲符有相同的發音部位，共可正確推測約 79% 的形聲字，再加上轉音規則的部分，共可將推測正確度提高到 82%。在第三層的發音關聯規則中，皆可以少數幾條規則揭露大部分形聲字發音的脈絡，由此幫助中文學習者概觀形聲字與其聲符的發音特性，提供他們簡單好記的發音規則。

另外，在這些主要的形聲字發音規則中，雖然大部分的形聲字發音與其聲符發音相同，但也有不少規則揭露例外的情形，稱為轉音規則，這些規則幫助學習者增加推測形聲字發音的正確性。故進一步探討在三層當中的轉音規則，第二層前五項轉音規則如表九所示。

進一步地，本研究實作一個網站呈現規則的視覺化，以此作為形聲字識字教學輔助系統。如圖五、圖六所示，藉由互動的方式，使用者可依照有興趣的項目，在發音階層圖中點選規則分類的方框，例如點選「舌尖前」，網站便呈現符合舌尖前分類的規則視覺化

圖。使用者也能進而使用下拉式選單查詢符合規則的形聲字，例如選擇”聲符=ㄆ→形聲字=ㄑ”，結果呈現符合的形聲字與其聲符發音，共 13 個常用形聲字與 37 個非常用形聲字。另外亦提供英文版的網站系統，將注音符號以漢語拼音的方式呈現給學習者。如此一來，漢語學習者透過視覺化的方式，能輕鬆記憶發音關聯式規則，並在操作識字教學輔助系統的同時，增加學習經驗與識字量，減少學習負擔。而此系統亦可輔助漢字研究，提供重要的發音規則予以參考與應用。

表 六、在不同層級上的發音規則數

層級	主要發音規則數
Level 1 發音層	275
Level 2 聲韻層	99
Level 3 聲韻分類層	34

表 七、主要形聲字聲母「發音部位」發音規則

Level 3-發音部位						
ID	聲符聲母	則	字聲母	普遍性(%)	正確性(%)	舉例 (聲符：字)
1	舌面	→	舌面	15	75	齊(ㄑ一2)：擠(ㄑ一3)
2	舌尖中	→	舌尖中	13	86	屯(ㄊㄨㄣˊ2)：頓(ㄊㄨㄣˊ4)
3	零聲母	→	零聲母	12	76	于(ㄩ1)：宇(ㄩ3)
4	舌尖後	→	舌尖後	11	71	專(ㄓㄨㄢˊ1)：傳(ㄓㄨㄢˊ4)
5	雙脣	→	雙脣	10	92	八(ㄉㄨㄚˋ1)：趴(ㄉㄨㄚˋ1)
6	舌根	→	舌根	10	79	鬼(ㄍㄨㄟˋ3)：塊(ㄍㄨㄟˋ4)
7	舌尖前	→	舌尖前	5	77	卒(ㄗㄨㄟˊ2)：翠(ㄗㄨㄟˋ4)
8	脣齒	→	脣齒	3	73	凡(ㄈㄢˊ2)：帆(ㄈㄢˊ2)
9	舌尖後	→	舌尖中	2	16	丑(ㄔㄨㄟˊ3)：妞(ㄔㄨㄟˊ1)
10	脣齒	→	雙脣	1	26	分(ㄈㄢˊ1)：扮(ㄈㄢˊ4)

表 八、前五主要形聲字「聲母」發音規則 (依普遍性排)

Level 2-聲母						
ID	聲符聲母	則	字聲母	普遍性(%)	正確性(%)	舉例 (聲符：字)
1	空	→	空	12	76	憂(ㄩㄞ1)：優(ㄩㄞ1)
2	ㄉ	→	ㄉ	6	93	良(ㄉㄞˊ2)：浪(ㄉㄞˊ4)
3	ㄌ	→	ㄌ	5	56	吉(ㄌ一2)：結(ㄌ一ㄝ2)
4	ㄒ	→	ㄒ	4	65	星(ㄒ一ㄥ1)：醒(ㄒ一ㄥ3)
5	ㄑ	→	ㄑ	4	54	者(ㄑㄜˊ3)：煮(ㄑㄜˊ3)

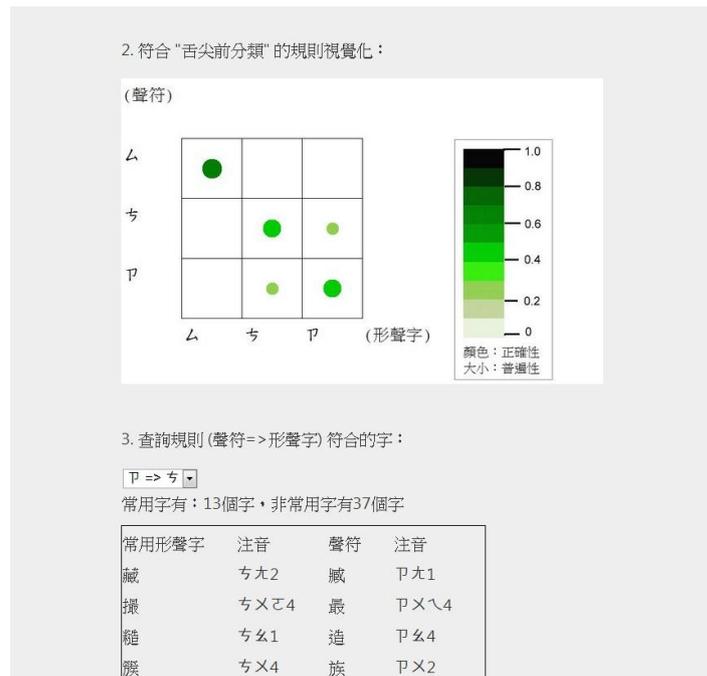


圖 六、注音符號版網站視覺化-規則視覺化

五、結論與未來方向

本研究結論可分三個部份，首先由影響形聲字發音的因素分析中，分析出最能影響形聲字發音的屬性就是聲符的發音，以此作為規則探勘的項目，幫助使用者更容易判斷形聲字的發音，並大幅減少探勘後的規則數目。第二部分為輔以漢語音韻學的知識，將漢字發音分成三個層次來看，建立漢字發音階層架構，進行多層次發音關聯規則探勘，篩選規則找出「主要的形聲字發音規則」，由此輔助漢語學習者與漢字研究歸納形聲字發音的情形、了解漢字發音的脈絡。再來。第三部分為設計視覺化的方法來呈現規則，使用者可一目了然規則的涵蓋範圍與其重要程度、易於學習發音規則，並將此以互動式的網站系統呈現，利於使用者選擇有興趣的規則分類，再輔以常用字與非常用字，讓漢語學習者增加學習經驗與識字量。期望能由本研究具體提供的漢字形音關係與組字特性，幫助學習者有系統的方式學習，減少學習負擔並增加識字能力，亦協助漢字研究有更進一步的發展。

在未來，本研究將會延續研究團隊的計畫，將研究成果與漢字單元教材結合，由單元課程中學到的生字，延伸學習發音規則與相對應的形聲字，並進行實地教學施策。而學習者需了解有關聲韻分類的「發音方法」、「發音部位」、「韻別」和「嘴型」，對於這項要求是否會造成學習者的負擔，希望能在未來實際教學中得到回饋並改進。此系統亦可以其他拼音法呈現漢字的發音，期望能幫助更多外國學習者加強他們的識字能力。另外，對於簡體字中的形聲字發音，是否依然與聲符發音存在著緊密的關聯，可在未來研究中進一步探討。未來仍朝向發揮漢語數位學習的優點，協助漢語學習者奠定好漢語基礎能力，並同時以生動易懂的方式呈現，不但可增加學習者的學習效率亦能引起他們的學習興趣。

參考文獻

- [1] 張良民, “全球華語學習熱潮與僑教發展”, *研習資訊*, 2006 年, 23:2, 9-15 頁。
- [2] 林季苗, “漢語教學四大原則與法國經驗”, *華語文教學研究*, 2011 年 8 月, 8:2, 65-79 頁, 。
- [3] 段玉裁《說文解字注》, 十一版, 黎明文化事業股份有限公司, 台北, 民國八十三年七月。
- [4] 費錦昌, “現代漢字部件探究”, *語言文字應用*, 語文出版社, 1996 年, 第 2 期總第 18 期, 20-26 頁。
- [5] 辜玉旻、柯華葳、高嘉慧, “識字教學法與口語詞彙能力對新移民女性中文識字學習之影響”, 中央大學學習與教學研究所碩士論文, 2010 年。
- [6] 高柏園、郭經華、胡映雪, 華語文作為第二語言之字詞教學模式與學習歷程研究, 2009-2010 年。
- [7] 國家語言文字工作委員會, *現代漢語通用字表*, 中華人民共和國新聞出版總署, 中國大陸, 1988 年。
- [8] 中研院文獻處理實驗室, “漢字構形資料庫”, [Online]. Available: <http://cdp.sinica.edu.tw/cdphanzi/>。
- [9] Lee, C.-Y., Tsai, J.-L., Su, E. C.-I., Tzeng, O. J.-L., & Hung, D. L., “Consistency, regularity and frequency effects in naming Chinese characters”, *Language and Linguistics*, 6(1), pp. 75-107, 2005.
- [10] 張嘉惠、李淑瑩、林書彥、黃嘉毅、陳志銘, “以最佳化及機率分佈判斷漢字聲符之研究”, *ROCLING*, 2010。
- [11] 張嘉惠、林書彥, “聲符部件排序與形聲字發音規則探勘”, *ROCLING*, 2011。
- [12] Jiawei H. and Micheline K., *Data Mining: Concepts and Techniques*, 2nd ed., Morgan Kaufmann Publishers, March 2006.
- [13] Michael Hahsler and Sudheer Chelluboina, “Visualizing association rules in hierarchical groups,” In *Computing Science and Statistics, Vol. 42, 42nd Symposium on the Interface: Statistical, Machine Learning and Visualization Algorithms (Interface 2011)*, the Interface Foundation of North America, June 2011.
- [14] Gupta, G, Strehl, A., and Ghosh, J., “Distance Based Clustering of Association Rules,” in *Intelligent Engineering Systems through Artificial Neural Networks (Proceedings of ANNIE 1999)*, 1999, pp. 759-764.
- [15] Wikipedia, “Mutual information”, available at: http://en.wikipedia.org/wiki/Mutual_information/ (accessed March 2013), 2013.