

對外漢語教學的數字化工具之研究

^{1,2}董名垂 ^{1,2}黃輝 ¹關鎮霆

³張九齡

¹澳門電腦與系統工程研究所

²澳門大學

³澳門理工學院

{dmc, derek}@inesc-macau.org.mo

摘要

隨著全球化和兩岸四地經濟的蓬勃發展，越來越多的外國公司希望同兩岸四地進行貿易往來，互通有無；越來越多的外國人希望學習中文、深入了解中華悠久的歷史和文化。然而，首當其衝把這些外國公司和外國人擋在博大精深的中華歷史和文化之外的就是中國複雜的語言文字，這對於那些來自拼音文字國家的人員而言，尤其如此。人們期盼能借助現代 IT 技術和電腦，提供學習工具，鋪就橋梁，減小阻隔。無疑，提供學習漢語之數字化工具是時代的迫切需要。

本文針對外國人學習漢語必須掌握漢語詞彙、拼音、聲調三要素，同時結合漢語對外教學中“以詞為本”的成功教學經驗，創新提出並實現了將漢語詞彙、拼音、聲調三要素捆綁一起，在電腦錄入和顯示時同時出現的“捆綁技術”，以及將漢語拼音、聲調與漢語詞彙分拆以方便進行編輯練習題的“鬆綁技術”。在此基礎上，進一步研究並實現了對整段中文一次性自動加註或分拆漢語拼音及聲調的功能。

為了令本數字化工具適合教與學兩方面人員使用，本文勾劃出擬增加漢語發音訓練和輔以圖像的多媒體教學設想，以及針對學員學習時出錯問題能智能化生成相關習題並自動統計練習成績的功能。最終，還擬將增加旨在幫助學員正確書寫漢字和校準漢語發音的更高智能化功能。

壹、IT 同漢語教學相結合的創新點與難點

早在七十年代中期，當基本上解決了漢字在英文電腦上的顯示及漢字的雙字節內碼表達技術以後，IT 技術就已經被引進了文字的編輯、排版、處理等項工作，並日益發揮出越來越大的作用。中國大陸的金山詞霸、東方快車、南極星、步步高，臺灣的文曲星，美國的好易通、快譯通等等電子字典以及著名的北大方正排版系統、激光照排系統等等如雨後春筍般紛紛推出；進入八、九十年代，許多研究開發工作集中在機助自動翻譯和語音辨識系統；而進入 21 世紀，更有人開始研究人機對答咨詢、視覺翻譯以及口語對口語直接翻譯等系統。然而，從對外漢語教學的角度來看，即使有了上述各類數字化系統，但至今仍然缺乏一套可用

於漢語教與學的電腦軟件工具，從而能幫助教師快速編輯漢語教材、習題和試題以及幫助外國學生高效學習漢語。

衆所周知，外國人學習漢語必須掌握漢語詞彙、拼音和聲調三要素，但是目前任何一種字處理軟件系統，例如像微軟的 Word、Notepad 等都不能在一行中同時鍵入或同時顯示這三要素。一個無奈的替代辦法就是先鍵入漢字，然後在另一行加註漢語拼音，最後再在漢語拼音上方採用加註符號的辦法添加四聲音標。其麻煩之大和出錯率之高，足以令大部分教師叫苦不迭和望而卻步。基於此，一個創新的想法應運而生，那就是採用 IT 技術研究開發漢語詞彙、漢語拼音和聲調的“捆綁技術”、“鬆綁技術”、“整段一次性自動加註和分拆技術”。

在研究“捆綁技術”中要解決的關鍵難點是：

- ❖ 捆綁漢語詞彙、其漢語拼音和聲調，從而可在一條水平行線上同步出現和編輯漢語詞彙、漢語拼音和聲調符號這三種不同字符；
- ❖ 對漢字中多音字、多聲調字提供漢語拼音全集供選擇；
- ❖ 整合各常見輸入法，並能在 Word、Notepad 等環境中使用；
- ❖ 有效解決漢語簡繁體之間的切換和內碼自動轉換；
- ❖ 對整段中文一次性自動加註漢語拼音及聲調；
- ❖ 對整段拼音+聲調一次性自動加註漢語詞彙。

在研究“鬆綁技術”中要解決的關鍵難點是：

- ❖ 分拆漢語詞彙、拼音+聲調，從而可將漢語詞彙、拼音+聲調分別獨立進行編輯操作；
- ❖ 對整段中文一次性自動分拆漢語詞彙、拼音+聲調。

在研究“整段一次性自動加註和分拆技術”中要解決的關鍵難點是：

- ❖ 有效解決整段中文文章裏詞的切分問題；
- ❖ 漢語詞彙的快速搜索、匹配及資料拾取；
- ❖ 漢語拼音+聲調的快速搜索、匹配及資料拾取。

貳、系統設計及其技術實現

在設計前的需求分析階段必須確定：選擇、錄入和處理的漢語對象是“以詞為本”，還是“以字為本”？經過對比研究得出結論：“以詞為本”便於學生加深對漢語的理解和記憶；以詞為本可以減少漢語中的多音字、多聲調字的選擇；以詞為本可解決“捆綁”時漢語拼音中的分詞連寫問題。所以，無論從方便語言教學、編寫教材和習題，還是從有利於學生學習、理解和記憶等幾個方面考慮，“以詞為本”作為本資料庫的輸入單位都大大優於“以字為本”，其不足之處是令建庫時的輸入工作量大大增加。

在總結了漢語普通話教學、對外漢語教學“以詞為本”的教學特點和規律之基礎上，制定出對軟件系統的功能要求和設計技術指標，完成了如下圖 1 所示的 MFHCSS (Multi Function HanCi Study System) 系統功能模塊設計：

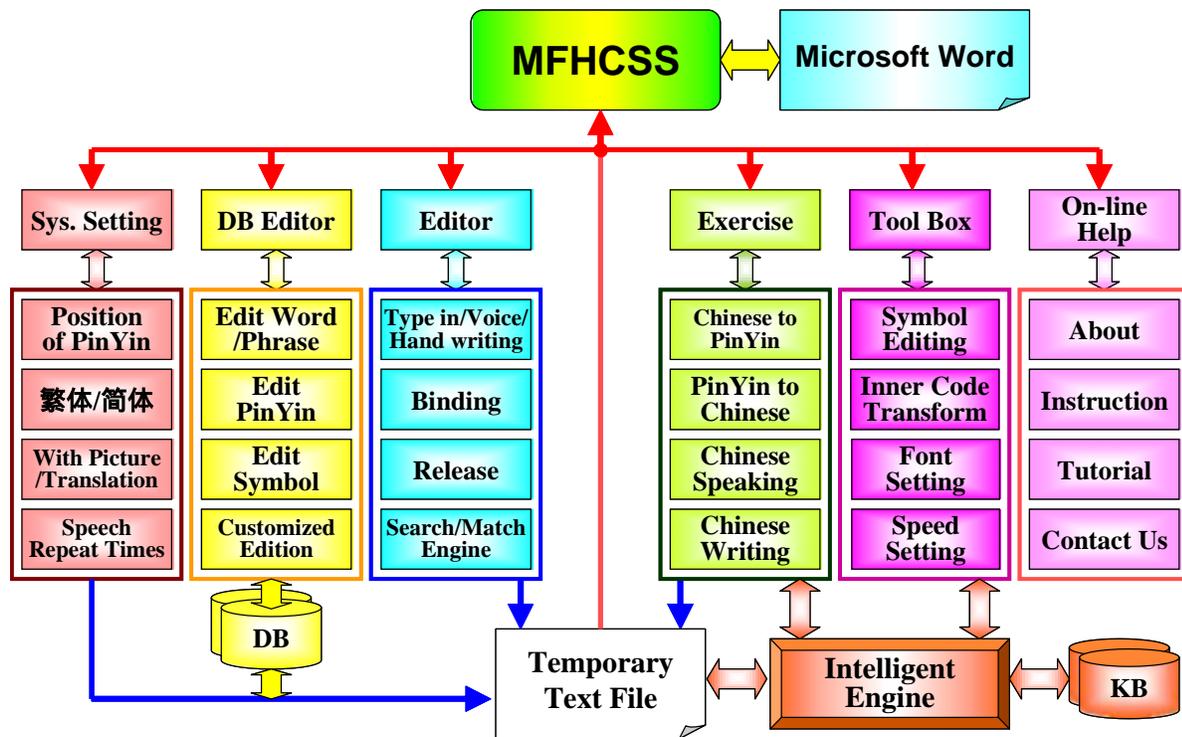


圖 1. MFHCSS 系統功能模塊圖

其中：

- ❖ 系統設置 (Sys. Setting)：設置系統參數及屏幕顯示參數，例如設置拼音+聲調的擺放位置、選擇簡繁體、顯示漢語詞組所對應的外語譯文、顯示釋意圖片、設定跟讀的重複次數等功能；
- ❖ 數據庫編輯器 (DB Editor)：前臺授權增刪詞庫內容，並可建立客戶化詞彙庫；後臺進行數據的搜索及匹配；
- ❖ 編輯器 (Editor)：實現數據輸入、捆綁、鬆綁、整段加注/分拆拼音+聲調、整段加注漢字等功能；
- ❖ 習題練習 (Exercise)：自動生成各類作業練習，例如由漢詞到拼音、由拼音到漢詞、中文發音訓練、書寫漢字訓練等，以及顯示對歷次作業結果的統計數據；
- ❖ 工具箱 (Tool Box)：提供必要的教學輔助功能，例如進行特殊符號編輯、漢字內碼轉換、漢字字體字形選擇、設定習題練習的速度等功能；
- ❖ 在線幫助 (Help)：教授如何操作本系統、提供幫助信息；

- ❖ 智能引擎 (Intelligent Engine)：鏈接知識庫及人工智能模塊，對習題練習、編輯器、工具箱等模塊注入人工智能。

在技術上，實現“捆綁技術”的基本要領是，預先錄入加註好拼音及聲調的漢語詞彙，建立漢語詞彙庫。在用戶調用本系統工作時，軟件系統將在線實時按漢語簡繁體設定以及內置的漢語切詞規則在漢語詞彙庫中進行詞彙或發聲的快速搜索及匹配。匹配上的，即刻拾取資料並把相應結果顯示在字處理工作環境中；匹配不上的，則保持原文不動。實現“鬆綁技術”的基本要領則是，軟件系統將把加註好拼音及聲調的漢語詞彙逐一進行詞彙同拼音+聲調的分拆，並把分拆開的拼音+聲調集總顯示在原文的上方。

因本系統將被用於漢語普通話教學和對外漢語教學，因此所選錄的漢語詞彙必須符合《普通話水平測試大綱》、《對外漢語教學大綱》及有關權威字典的要求和標準。我們從《普通話水平測試大綱》、《對外漢語教學大綱》、《現代漢語規範詞彙》、《現代漢語規範成語》、《現代漢語規範慣用語》、《世界地理》和《中國地理》等資料中選擇全部或部分常用詞彙、專業詞彙，最終在詞彙庫中錄入了 14 萬多條漢語詞彙，並且完全按《中華人民共和國國家標準 ICS 01.140.10A GB/T 16159----1996》“漢語拼音正詞法基本規則”嚴格進行錄入。

隨著詞彙庫內容的不斷豐富和加大，數據在詞彙庫中的搜索及匹配的快速性與準確性將成為一大瓶頸問題。通過文字分類處理技術、建立決策樹和元知識的編寫等智能化技術，本系統基本解決了這個難題。

此外，因為圖 1 中各功能模塊所要實現的子功能各不相同，為了提高各功能模塊的運行效率，所以開發時充分結合它們各自的功能特點，有針對性地選用了那些最適宜的編程語言來分別實施開發。例如採用 VB 開發用戶界面，以實現系統管理及編輯器等功能；採用 C++ 開發習題練習、發音訓練、工具箱等功能；採用 FoxPro 實現數據庫功能及操作；採用 CLIPS 及 C++ 實現人工智能模塊等。最後通過接口集成技術，將各功能模塊整合成一個完整的系統，再把通用漢字輸入法集成到本系統，並且最終把該系統嵌入到微軟公司的 Word 字處理系統中。

叁、系統操作及功能

在操作使用本系統時，用戶通過調用主頁鍵入漢語拼音或採用任何通用的漢字輸入法鍵入漢字，該主頁就會接受鍵入的數據，經智能化判別後自動傳遞到數據搜索及匹配引擎在相應的詞庫內進行符號或詞彙的搜索和匹配，下圖 2 顯示出了這一過程。一當匹配完成，其結果就會被迅速傳遞到主頁並以“捆綁”的方式將漢語詞彙、拼音及聲調同時顯示在電腦屏幕上或者直接顯示在微軟的 Word 頁面內，以供進行編輯修改。因為在本系統中整合了通用的漢字輸入系統，所以同樣可採用拼音、五筆、倉頡等輸入法並帶有智能化聯想功能。因此用戶可僅僅鍵入漢語詞彙之拼音的幾個首字符，系統就能聯想搜索到相應的漢語詞彙。其後按照用戶的需要，可調用“鬆綁”功能將捆綁顯示在一起的漢語詞彙、拼音+聲調分拆成單獨的漢語詞彙和拼音+聲調，這些分拆後的漢語詞彙、拼音+聲調可被分開獨立進行編輯和使用。本系統還提供了可“整段加注拼音及聲調”或“整段加注漢語詞彙”的功能。用戶可選定一長段漢語文字，然後點擊“整段加注拼音及聲調”的功能圖標，於是系統可將所有對應的拼音及聲調加注到相應的長段漢語原文之上。反之亦然，用戶選定一長段漢語拼音及聲

調，然後點擊“整段加注漢語詞彙”的功能圖標，於是系統可將所有對應的漢語詞彙加注到相應的長段拼音及聲調之下。在今後的升級版本中還將加入從語音到文字的輸入功能，從而令用戶可通過話筒和說話來輸入漢語字詞及其拼音+聲調。

在調用習題練習功能模塊的時候，用戶可選擇“由拼音到漢字”、“由漢字到拼音”、“發音訓練”等練習。系統將從詞庫中隨機選擇十多個漢語詞彙並用這些詞彙自動產生有關的練習題並記載學員的應答情況，最終作出應答統計並顯示統計結果。在今後的升級版本中還將增加智能化記憶功能，令本系統能自動記憶用戶常犯的錯誤，從而在進一步出題時能針對他常犯的錯誤增多有關的習題練習。在調用“發音訓練”功能模塊的時候，用戶點擊漢語詞彙，系統將用內置語音合成器發出該詞彙的普通話發音，每次點擊後的發音次數可在系統管理模塊中設定。在今後的升級版本中還可考慮增加語音辨識功能，從而可記載學員的跟讀聲音並同標準的發音相比較，最終顯示出該學員發音的正確度百分率。此外，還將增加一筆一劃教授如何書寫漢字的更高智能功能。

調用系統設置功能模塊，用戶可以設置不同的系統參數或屏幕顯示參數，例如設置每次點擊漢語詞彙以後發音幾次，將拼音加聲調顯示在漢字的上/下/左/右方，選擇漢字及拼音的字體字形，是否添加特殊字符或進行特殊編輯等。

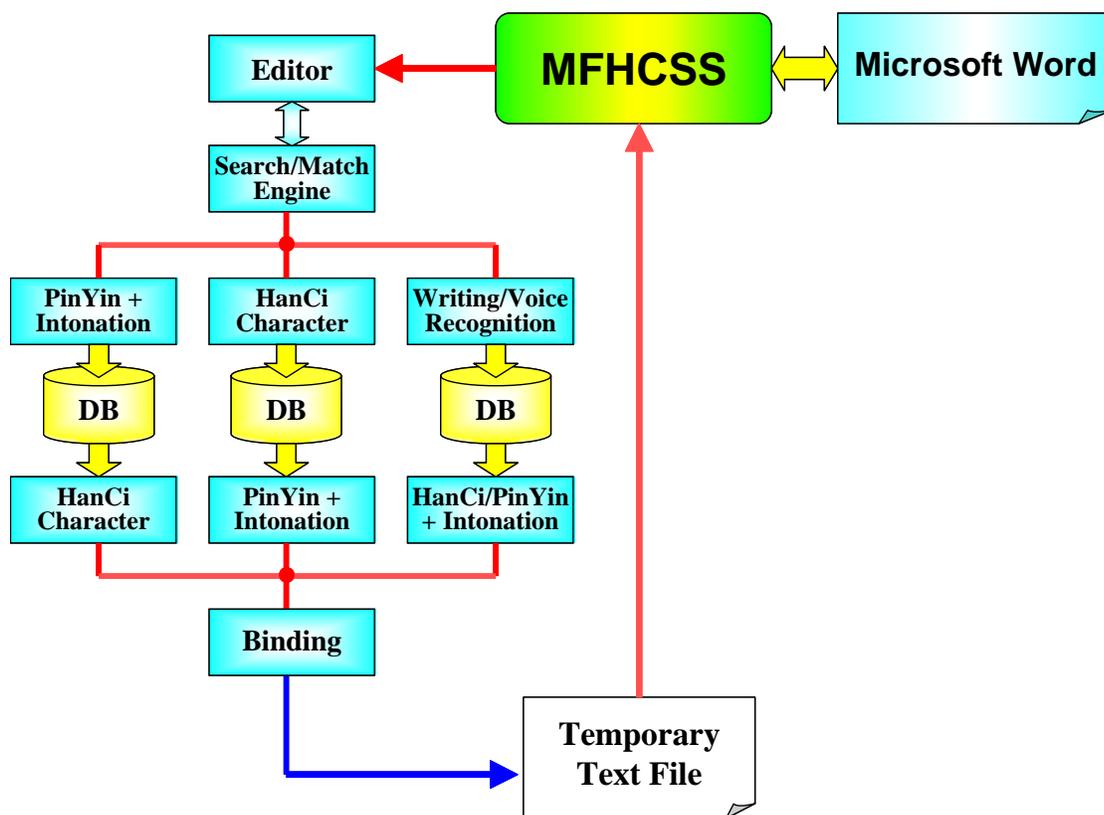


圖 2. 漢語詞彙或拼音+聲調的搜索及匹配功能示意圖

肆、 結束語

將 IT 技術同漢語教學結合，開發出方便實用的數字化工具，不僅為從事漢語教學的老師們提供了編寫教材、出練習題及考卷的便捷，同時也為外國人學習漢語提供了教與學的一種符合時代潮流的高效手段。除了前述的漢語發音訓練、書寫漢字訓練，本系統通過調用工具箱功能模塊還可設置顯示外語譯文及釋意圖片。這樣，用戶在看到捆綁式顯示的漢語詞彙、拼音及聲調的同時，還可看到該漢語詞彙的外文解釋以及相關的釋意圖片。按照所使用外語的不同，今後的系統將分為英語版、法語版、葡語版等等。

總之，採用 IT 技術及人工智能技術，結合漢語教學的特點，本系統“以詞為本”，擬提供集“捆綁顯示”、“鬆綁編輯”、“整段加註”、“整段分拆”、“導讀發音”、“教授書寫”、“外文解釋”、“圖片釋意”、“習題練習”等多功能於一體的對外漢語教學軟件，在應用資訊科技於漢語教學上進行了研究及探索。初步的成功表明，這一方向是正確的，值得投入力量繼續做更有成效的研究開發。

伍、 參考文獻

- [1]刘颖，《计算语言学》，北京清华大学出版社，2002
- [2]荀恩东、赵铁军、李生，《英汉机器翻译中译文选择方法》，《情报学报》，1999
- [3]栾浩，《基于合一的汉语句法分析系统的分析和实现》，(硕士学位论文)，北京清华大学计算机系，1993
- [4]Schneider G, Dowdall J and Rinaldi F, “A robust and hybrid deep-linguistic theory applied to deep-linguistic parsing”, Proceedings of Robust Methods in the analysis of NL data (ROMAND) workshop, Coling-2004, Switzerland Geneva, 2004
- [5]Wilks Y A, “A Preferential Pattern-Seeking Semantic for Natural Language Inference”, Artificial Intelligence, 1975