

漢字知識本體在漢字教學的應用*

周亞民
銘傳大學

摘要

漢字不僅是字，也經常是構詞的語素，在運用和教學中都是非常重要的基礎。在電腦的發展中，漢字的數位化也相當的重要，它不僅是中文電腦的基本要件，漢字的數位化更能夠作為學習漢字的工具。過去，已經建立了很多漢字數位學習工具，近年來，多媒體和網路科技的發展，讓漢字學習的方法更多元化，但是其內容深度仍有待改進。這些問題包括：普遍缺乏漢字歷時變異的知識、缺乏漢字關係的知識、過度簡化正簡漢字的關係、未能與其它語言資源連接、未採用有分享能力的形式語言等。本論文旨在說明這些問題在華語文學習的角色，以及漢字知識本體（Hantology）如何解決這些問題，並提出利用漢字知識本體作為語言資源，增加電腦處理正體中文的能力，提升正體字資訊在網路的優勢，增加學習正體字的需求。

關鍵詞：書寫系統，華語文教學，數位學習，語言知識本體

1. 緒論

漢字數位學習工具能夠讓學習者獲得哪些漢字知識，主要取決於被數位化的漢字知識，尤其在沒有教師而由學習者獨自使用數位教材的情境，因此漢字數位教材提供的內容，對於學習者的漢字學習歷程，扮演關鍵的角色。本研究分析漢字數位化的情況，探討目前數位學習工具缺乏的漢字知識，並以漢字知識本體為例，說明這些缺乏的漢字知識如何表達在計算機，漢字知識本體可以讓學習者獲得哪些漢字知識，並提出對漢字數位教材發展方向的想法。

回顧過去漢字數位化的發展，字形的數位化是最早受到重視的問題，為

* 誌謝：本研究感謝國科會人文處提供的支持（計劃編號： NSC 95-2411-H-228-001 和 NSC- 96- 2411- H- 228- 002- MY2），以及本刊審查人提出的修改建議。

了能夠呈現漢字的字形，一個漢字字形使用一個字碼，形成編碼系統後，運用點陣圖或向量圖與字碼對應後，將字形呈現給使用者。計算機常用的漢字編碼系統包括：BIG-5、Unicode、CNS 11643¹等編碼系統，但是這個方法對於漢字的教學和中文資訊處理都是不足的，因為硬體上的系統字形是無法滿足文件在美觀上的要求，因此，部份軟體公司將硬體的系統字形與編碼分開，增加其它漢字字體，以字體檔案的方式提供使用者，將編碼與不同的字體檔案對應來呈現不同的字體，例如華康和文鼎等字體。但這些字體不僅缺少古文字，數量也不足，因此無法滿足文字學相關學者，以及古籍數位化的需要，因此，學者們爲了研究上的需要，自行建立了甲骨文、金文、小篆等字體和資料庫，以方便利用計算機整理漢字字形、古籍數位化和發表研究成果（謝清俊 1992）（莊德明、謝清俊 2005）（黃沛榮 2001）（陳昭容 2004）。

隨著計算機在多媒體資訊處理能力的提昇，漢字的字音也被數位化，以作爲漢字教學工具或是與計算機互動的介面（黃大一 1988）（石鴻珍 2006）。字音數位化的表達方式可分爲兩類，第一類是以拼音符號表達漢字的拼音方式，另一類是使用數位音樂的檔案格式，優點是讓使用者直接聽到實際的發音，有利於學習與糾正發音。利用計算機輔助漢字的教學，除了可以學習字形和字音，還可以學習筆順和字義。爲了讓華語文學習者能夠看到筆劃書寫的順序，習得漢字筆順，筆順需以動畫呈現（趙慧如 1991）。漢字字義的部份則是以文字的方式表達，針對華語不是其母語的學習者，除了中文的解釋，另外增加英語的詞語解釋。

到目前爲止，漢字被數位化的內容，已經包含了字形、筆劃、字音和字義等知識，因此，學習者可以藉由漢字數位學習工具獲得這些知識，漢字數位學習工具彼此之間的差異不大，多數都能夠支援字形、筆劃、字音和字義的學習，不同之處主要是呈現方式，以及與學習者之間的互動方式。

2. 漢字數位化的問題

漢字數位化經過了二十多年的發展，漢字的字形、字音和字義等知識皆已被表達在計算機，但是，仍然有許多漢字的知識未被數位化或尚待改進，包括：字音和字義的變化、正簡字的關係、異體字的關係、漢字在計算機的知識表達方式與其它語言數位資源的銜接等問題，在過去的相關文獻中，缺

¹ CNS 11643 是國家中文標準交換碼，其編碼方式，請參考行政院主計處電子處理資料中心的全字庫網站，網址是 www.cns11643.gov.tw/AIDB/encodings.do#encode1

乏由此觀點進行的研究。

本研究將這些問題進一步說明如下：

2.1 缺乏字音和字義在不同時代的變化

多數的漢字數位教學工具，皆能夠提供使用者學習漢字字形、筆順、字音和字義，但是大多數提供當代華語的字音和字義，如果使用者需要學習古漢語和中古漢語的字音和字義，這些數位教學工具並未提供這方面的知識給使用者學習。

少數能夠提供字形或字音變化知識的工具，都是漢字資料庫，但是，這些資料庫大部份針對特定的漢字屬性，例如有些目標是建立字形變化資料庫，有些則是建立字音變化的資料庫（古宇時雄等 1999）（莊德明、謝清俊 2005）（黃沛榮 2001）（楊秀芳 2003）。這些資料庫大部份都沒有字義的知識，也沒有字義變化的知識，使用者要學習字形與字音變化時，必需要分別使用不同的資料庫，如果要學習字義的變化，則要依賴紙本的參考資料。

2.2 過度簡化正簡字之間的關係

多數漢字數位學習工具和漢字資料庫，將正簡字的關係簡化為對照表，但是只用對照表，無法完整呈現正簡字的關係，也難以讓使用者學會正簡字的使用差異。中國大陸漢字簡化方案所公布的簡體字，有些在古漢語或中古漢語就已經使用，例如有些簡體字以前就曾經是俗字，只是後來用法差異擴大，而成為兩個不同的漢字，這些關係在許多漢字數位教學工具之中，並沒有被適當地說明。

2.3 異體字關係的表達仍需改進

漢字的使用已經有數千年的時間，在這麼長的時間中，字形、字音、字義和異體字皆有變化，而形成複雜的漢字關係。例如古今字、通假字、正俗字皆是漢字不同的關係，訓詁學者以同義或近義字訓釋，或以反義字反訓，皆指出了漢字之間的同義和反義關係。這些關係皆是文字學家重要的研究重心，但是，多數的漢字數位學習工具和資料庫，皆缺乏漢字關係的描述，主要原因是同時要描述字形、字音、字義和異體字的變化和關係，其複雜度大幅增加，因此，這些知識並沒有被適當表達在計算機，使用者也難以獲得這些知識。

根據 WordNet 的研究成果，由於它在計算機表達了英文詞彙的關係，對

於學習英語的學習者，以及檢索和自然語言處理都有很大的幫助（Miller 1995）。因此，若能夠在漢字數位學習工具和漢字資料庫中，建立漢字彼此的關係，對於學習漢字或計算機進行中文資訊處理將會非常有幫助。

與拼音文字相較，漢字的異體字關係最具特殊性，因為漢字字形是概念的具體表徵，由於漢字並不是特定的人所創造，不同字形表示相同的概念，或不同的概念使用相同的字形，都是很普遍的現象，依據使用時間和地區的不同，字形意義也發生了變化，最後交織成複雜的一詞多形關係，將不同的字形連結成為異體字。異體字雖然是漢字的重要特性，但是只有少數的漢字資料庫和電子字典有異體字，例如：漢字構形資料庫²、教育部異體字字典網路版³、電腦漢字字形與異體字詞彙整合知識庫⁴。這些資料庫對異體字的知識表達，突破了很多的困難，但是對異體字之間的字形、字音和字義的關係較為簡略，不容易讓使用者清楚的了解異體字間的關係。

2.4 未採用有分享能力的形式語言

多數的漢字數位教學工具和資料庫，都是設計給使用者查閱而不是讓計算機用來處理中文，因為其數位化的表達方式，並沒有採用有分享能力的形式語言（formal languages），這樣的表達方法難以讓其它的應用系統分享漢字的知識。在學習知識的過程中，分享相當的重要，因為藉由分享可以匯集知識，進一步產生新知識，這就是語意網路（semantic web）的精神，因為網路上雖然有大量的資源，若很多資源的知識表達都未採用形式語言，就無法為其它計算機所理解和分享（Berners-Lee et al. 2001）。

漢字知識是否能夠被不同資訊系統分享，是非常的重要問題，因為漢字知識是計算機處理中文所需的基本知識，必須讓不同的中文資訊系統，都能獲取漢字的知識，這就需要使用語意網路的形式語言，但是目前沒有數位教學工具與資料庫使用形式語言。除了缺乏應用語意網路的形式語言，也只有

² 由中研院資訊所文獻處理實驗室謝清俊教授和莊德明先生等開發，收錄古今漢字 116,790 個，異體字 12,208 組，它是提供漢字字形最重要的資料庫之一，主要貢獻在於古今字形的數位化、字形結構與構字式分析。漢字構形資料可以提供網路下載，網址為 cdp.sinica.edu.tw

³ 民國八十九年六月由教育部完成正式版《異體字字典》，並多次改版，在異體字整理相當有貢獻，此部字典無論在文字學史上或字典編輯史上都當具有特殊意義。

⁴ 電腦漢字字形與詞彙整合知識庫是數位典藏國家型計劃技術分項計劃－建立 Unicode 漢字異體字表與異體字辭典之相關研究的成果之一，其研究目的是對 Unicode 的中日韓異體字進行整理，也是其主要的貢獻。

少數的漢字數位教學工具使用 SCORM (Sharable Content Object Reference Model)，作為與其它數位學習工具交換和整合的橋樑。若漢字數位教學工具採用 SCORM 標準⁵，易於整合不同的漢字數位教學工具內容，減少開發的時間，也可以與其它數位學習工具整合，增加教學內容的豐富度和活潑性。

2.5 未能與其它語言資源連接

由於網際網路的普及化，多語環境無論在真實或虛擬世界都已逐漸成形，因此，跨語言資訊檢索的應用越來越普遍，例如運用華語的關鍵詞，即可同時找到華語、英語和法語等不同語言的文件。為了要提供跨語言檢索，計算機必須利用雙語或多語的語言資源，透過共同的介面，將不同語言資源連接在一起，以完成跨語言資訊檢索。

實際上，如果漢字數位學習工具設計得宜，不僅可以用來教學，也可以作為網路的語言資源，讓計算機進行跨語言檢索時，提供所需要的語言知識。漢字數位學習工具若要作為語言資源，除了應使用語意網路的形式語言或 SCORM，還必需與其它語言的資源連接，才能在提供跨語言的資訊檢索。但是，目前漢字數位學習工具不僅未採用語意網路的形式語言，更缺乏與其它語言資源的連接，難以成為跨語言資訊檢索的資源。

漢字數位學習工具與其它語言資源連接，除了能夠作為跨語言資訊檢索的資源，對於母語不是華語的學習者，也有很大的幫助。如果要設計一個漢字數位學習工具，提供給以母語為英語的學習者使用，在字形、字音、字義加入英語的解釋是必要的設計，但是對於只會西班牙語不會使用英語的學習者，只有英文解釋，並不能提供良好的學習環境，因此，必需要再增加西班牙文的解釋，然而，要建立一個多國語言的漢字數位學習工具，非常的費時，若能與其它的語言資源連接，很容易建立多語的漢字數位學習工具，以滿足不同母語的華語文學習者。

為了呈現漢字數位化的現況，本研究將漢字數位化的知識加以分類，包括：部首、意符（形符）、聲符、字形、字形演變、字音、多媒體字音、字義、字義演變、異體字、異體字關係、衍生詞，並將數位化的形式表達方式和銜

⁵ SCORM (Sharable Content Object Reference Model) 是為美國 ADL 組織 (Advanced Distributed Learning) 規範之數位學習標準，詳細的規範請參閱：www.adl.org，或參考廖瑋榕、曾金金(2005)，〈符合 SCORM 標準的華語語音數位教學設計〉，全球華文網路教育研討論文集，僑務委員會主編。臺北：世界華語文教育學會。

接其它語言資源，呈現漢字數位教材和資料庫對漢字數位化的全貌（表一）。

表一、漢字數位化資源分析表

（符號說明：O 表示已數位化 V：表示已做數位化且為其特點）

	部首	意符	聲符	字形	筆順	字形演變	字音	多媒體語音	中古音、古音	字義	字義演變	異體字	異體字關係	衍生詞	銜接其它語言資源	採用 SCORM	使用語意網語言
IQChinese Go				O			O	O		O				O		O	
康軒電子教科書	O			O	O		O	O		O				O			
eStroke				O	V		O	O		O				O			
Live Interactive Chinese	O			O	O		O	O		O				O		O	
五百字說華語				O	O		O	O		O				O			
易學中文網路學校				O	O		O	O		O				O			
今昔文字鏡				V													
電腦漢字字形與詞彙整合知識庫				O			O					V		O			
漢字構形資料庫	O			V		V						O					
異體字字典網路版	O			O						O		V	O				
教育部常用國字標準字體筆順學習網	O			O	V		O			O				O			
教育部國語小字典	O			O			O			O				O			
漢字知識本體		V	O	O		O	O	O	O	O	V	O	V	O	V		V

3. 漢字知識本體簡介

漢字知識本體 (Hantology) 為表達漢字書寫系統的知識本體 (ontology)，它表達的漢字知識包括：意符、字形變異、字音變異、字義變異、異體字關係、詞彙衍生等 (Chou and Huang 2007)。漢字書寫系統的重要特性是書寫形式和字義是意符的衍生，因此，意符表達的概念是漢語書寫系統的核心。漢字知識本體以說文的五百四十部首作為漢字的基本意符，分析部首作為意符時所表達的概念，運用 SUMO (Suggested Upper Merged Ontology) 加以對應 (Niles and Pease 2003)，讓計算機不僅可以處理意符表達的概念和關係，還可以建立意符之間的關係和呈現漢字意符的知識結構，又能與 WordNet、中央研究院中英雙語詞網 (The Academia Sinica Bilingual Ontological WordNet) 結合 (Huang, Chang and Lee 2004)。

漢字知識本體將字義區分為本義、引申和假借義，其中本義以說文的釋義為依據，並採用 SUMO 表達字義的概念和關係。漢字知識本體描述了字音與字義關係和變化，以及不同字義的衍生詞彙，表達漢字的詞彙衍生關係。漢字知識本體表達了異體字的語境 (context)，語境的變數包括：字音、字義、聲韻、構詞和時間，用來描述不同的漢字在什麼語境情況可以交替使用。為了讓漢字的知識更容易於被分享和利用，漢字知識本體使用語意網路知識本體語言 (OWL-DL) 描述，並且與通用語言描述知識本體 (General Ontology for Linguistic Description) 整合，可提供計算機在自然語言處理所需的書寫、構詞和語法知識。

目前漢字知識本體已完成一千八百五十七個中文電腦最常用字的知識，其字頻為百分之九十七點三⁶，每個字表達的知識皆包括意符、字形變異、字音變異、字義變異、異體字關係和詞彙衍生，以及與 SUMO 知識本體的對應。由於漢字知識體描述了漢字之間的關係，尤其是異體字關係，若將異體字包含在內，漢字知識本體總字數超過一萬餘字⁷。

4. 漢字知識本體在漢字教學的應用

本論文提出的漢字數位化問題，包括了普遍缺乏漢字歷時變異的知識、

⁶ 依據林樹先生於交通大學所做的中文電腦常用字統計，1857 個最常用字已可涵蓋 97.3% 的使用頻率。

⁷ 漢字知識本體正進行次常用字的分析，預計在 2009 年 10 月可以完成三千九百二十二個漢字分析，若包含異體字，則超過一萬個漢字的分析。待完成內容二次校對，將提供學術界使用。

缺乏漢字關係的知識、過度簡化正簡漢字的關係、未能與其它語言資源連接、未採用有分享能力的形式語言等。這些問題的重要性，在於許多數位學習是使用者自主學習，並沒有老師的指導，因此，數位教材之內容是使用者學習知識的主要來源。本節進一步說明這些問題在華語文教學的角色和重要性，並以漢字知識本體說明如何表達這些數位教材所缺乏的這些漢字教學應用的優點，這些優點可以解決上述問題。

漢字知識本體在漢字教學應用的優點，包括：(1) 可以呈現漢字歷時變異知識，學習漢字於不同時代的字音和字義變化(2) 可以學習異體字關係(3) 可以學習正簡字關係(4) 提供對比分析的學習(5) 可以作為網路的語言資源，增加正體字的競爭優勢(6) 與其它語言資源連接(7) 採用有分享能力的形式語言。這些優點及其華語教學的角色和重要性說明如下：

4.1 提供漢字使用變化的知識

隨著中國大陸的崛起，學習簡體字越來越普遍，許多的文獻不斷強調學習正體漢字後，很容易就可以再學習簡體字，忽略了越來越多華語學習者已經學習了簡體字。實際上，我們面對的情況，不再是沒有學過華語文的學習者，而是對於已經使用簡體字者，如何設計一個能夠滿足和吸引簡體字使用者的數位學習系統。若是以華語文教學的市場角度來看，過去皆以尚未學習漢字和漢語的學習者為對象，現在，我們還要努力爭取已經使用簡體字的學習者。為了爭取已經會使用簡體字的學習者，正體字的教材必需強調差異化，不再只是字形、筆順、現代的讀音和現代字義等數位內容，這些是目前簡體字和正體字數位教學工具都應具有的功能，但是這兩者皆以斷代面的教學為主，並不重視漢字在不同時代的使用變化。

對於正體漢字的教學，不能採用與簡體字相同的內容，應該增加正體漢字的歷時面教學的比重，因為，相較於簡體字，正體字確實有很多優點，這些優點中以能夠幫助理解古籍的內容，並藉此更了解中華文化，最能突顯正體字優勢，也可能是簡體漢字的學習者，想要進一步學習正體字最務實的因素。因為若以中國大陸和臺灣的市場相較，學習簡體字是比較有吸引力，但是如果與中國大陸人民往來越多，文化差異將是必需要克服的問題。如果要了解中華文化，透過學習正體漢字意符、聲符、字形演變、字音演變、字義演變、正簡字的關係和異體字的關係，應該是很好的方法，然而很多漢字數位學習工具並沒有提供這些漢字知識。

部份數位學習的工具能夠提供漢字字形演變的知識（莊德明、謝清俊 2005；黃沛榮 2007, 2005），然而除了字形之外，漢字的字音和字義也發生了改變，但大部份的數位學習工具缺乏字音和字義演變的知識，而這些演變對理解古籍內容，非常的重要，因為當代漢語的字音和字義與古代漢語有許多差異，造成理解古籍內容的障礙。漢字知識本體對於華語文學習者，不僅能夠提供字形的演變，還能夠提供字音和字義的變化，將字音區分為古音、中古音和現代音，而字義的變化則區分本義、引申義和假借義，以及現代常用義，將古漢語與現代漢語的詞義加以區分，對於有興趣學習中華文化的學習者，可以幫助他們建立閱讀中國古籍的基本知識。



圖一、漢字知識本體描述「樂」的知識內容

圖一呈現的是「樂」在漢字知識本體的內容，漢字知識本體描述了其主要意符為「木」，字形演變包括了金文和小篆字形，其字音的古音韻部為「藥部」（依漢語大字典的古韻分類），中古音為「盧各切，入鐸來」，現代音是「ㄌ

ㄘ、」。「樂」的本義為「音樂」，說文的釋義為「五聲八音總名」，之後引申為「喜悅」和「娛也」，現代的常用義有「喜悅」、「喜歡去做」、「和諧動人的聲音」與「喜好」。

4.2 可以作為網路的語言資源，增加正體字的競爭優勢

漢字知識本體不僅能夠學習漢字在不同時代的使用情形，也運用語意網路的形式語言 OWL，使得漢字知識本體的知識，可以被計算機理解，而成為語言資源。在計算機擁有豐富語言資源的語言，該語言在未來資訊時代將有更好的競爭優勢，因為計算機能夠提供該語言使用者較好的資訊服務，滿足其資訊需求。英語是目前在計算機擁有最豐富語言資源的語言，尤其是 WordNet 語言資源的建立，更強化了英語在網路上的優勢，並引起歐洲國家對於英語在網路的強勢，深感憂慮。因為以英語表達的資訊，可運用的語言資源較多，在網路上能夠得到較好的處理，更加深了美國文化的擴散和影響力。歐洲各國有鑑於此，也建立了 EuroWordNet，並且已完成荷蘭語、義大利語、西班牙語等二十種以上的語言資源（Vossen 1998）。

目前，臺灣在計算機建立漢字資源仍然不足，這對於正體字的生存空間，是一大隱憂。近年來，不僅簡體字書籍的出版急速成長，在網際網路上，簡體字的資訊量也大幅增加，如果中國大陸逐漸解除網際網路的限制，其成長的幅度將更加提高。由於簡體字資訊在傳統媒體和數位媒體同時大幅成長，簡體字的擴散和影響力也隨之提高，雖然正體漢字擁有簡體字所沒有的優勢，但是如果在傳統媒體缺乏正體字的優良文學作品，在數位媒體上也缺乏良好的正體字語言資源，將減弱正體字的擴散和影響力，同時也會減少學習正體字的需求，那麼，即使正體字再好，也無法增加學習正體字的需求。

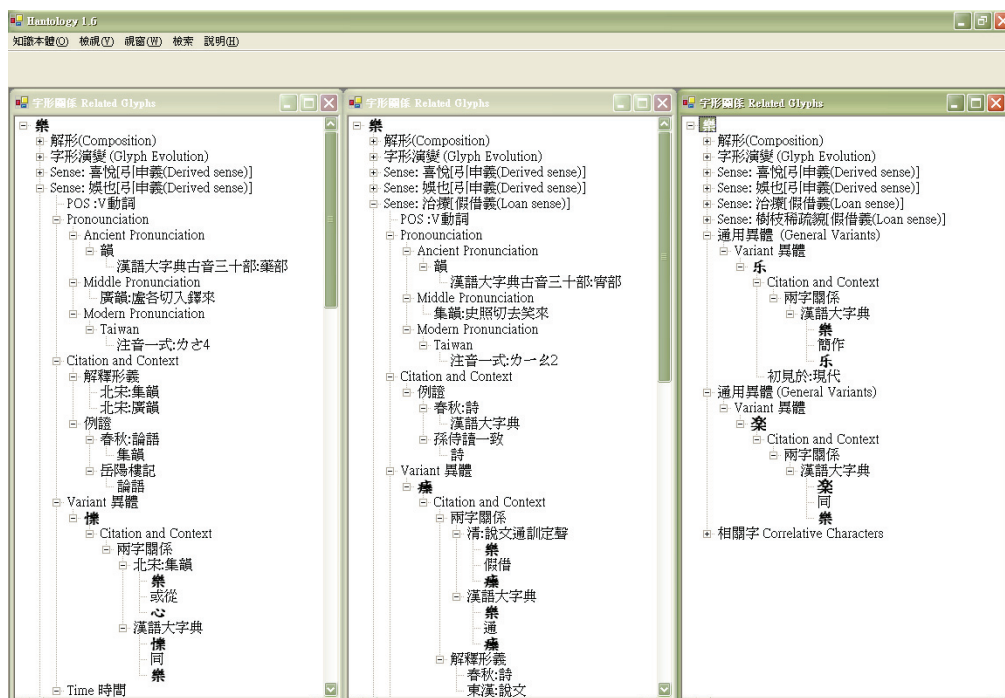
在建立漢字數位學習工具的同時，也應該要考慮此工具是否能夠作為計算機與網路的語言資源，如此一來，可以強化正體字的競爭力，而增加學習正體字的使用者，正體數位教材也才有足夠的市場。計算機應用在中文資訊處理已超過二十年，即使到現在，最基本的中文筆劃和部首排序都還是問題，缺字交換的問題也經常發生，尤其在人名、專有名詞、古籍和佛經的中文資訊處理，其中的原因之一，就是計算機缺乏漢字知識，而這些知識必需要由漢字的語言資源提供，但因缺乏漢字的語言資源，這些問題依然存在。

漢字知識本體不僅可以用來作為漢字數位學習工具，也是少數能夠作為語言資源的數位學習工具，一個主要的應用就是解決中文缺字問題。長久以

來，使用計算機處理中文，不可避免會遇到缺字，主要原因就是計算機缺乏漢字的語言資源，利用漢字知識本體提供缺字交換時所需要資訊，這個問題已經可以利用漢字知識本體解決（周亞民、黃居仁 2006）。漢字知識本體也能夠提供跨時間資訊檢索時，所需要的異體字知識，此功能對於古籍數位化和資訊檢索最適合，也更能發揮正體字在處理橫跨不同朝代文件的優勢。

4.3 可以學習異體字關係和正簡字關係

異體字是漢字的重要特性，異體字關係在多數的漢字數位學習工具都沒有被表達清楚，而漢字知識本體可以將異體關係，依字音、字義、時間和構詞等四度空間，描述兩個漢字在什麼時間、什麼字音和什麼字義時有異體字關係，以及其構詞的限制。

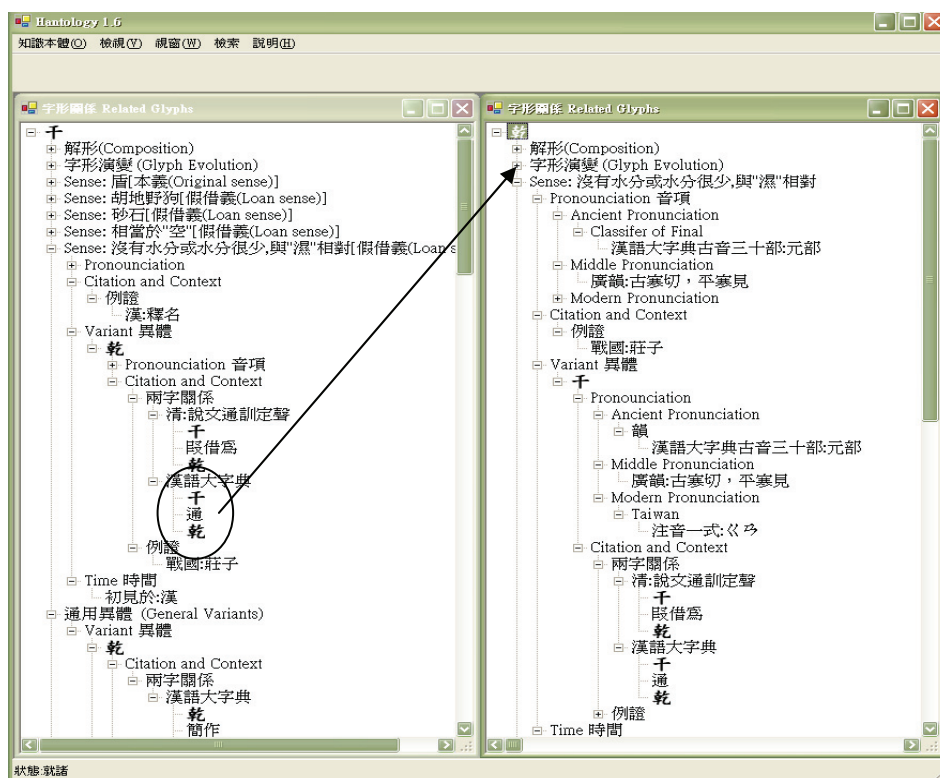


圖二、漢字知識本體描述「樂」的異體字關係

圖二以正體字「樂」為例，漢字知識本體描述「樂」在字義為「娛也」，中古音為「盧各切、入鐸來」，現代音是「ㄌㄜˊ」，「樂」與「𪛗」為異體字，在北宋的集韻以前即有此使用方法。當「樂」在字義為「治療」時，中古音

爲「史照切、去笑來」，現代音是「ㄌㄞˋ」，「樂」與「藥」爲異體字。而「樂」與「乐」、「樂」則爲通用異體字，不分字音和字義，兩者都可以交換使用。

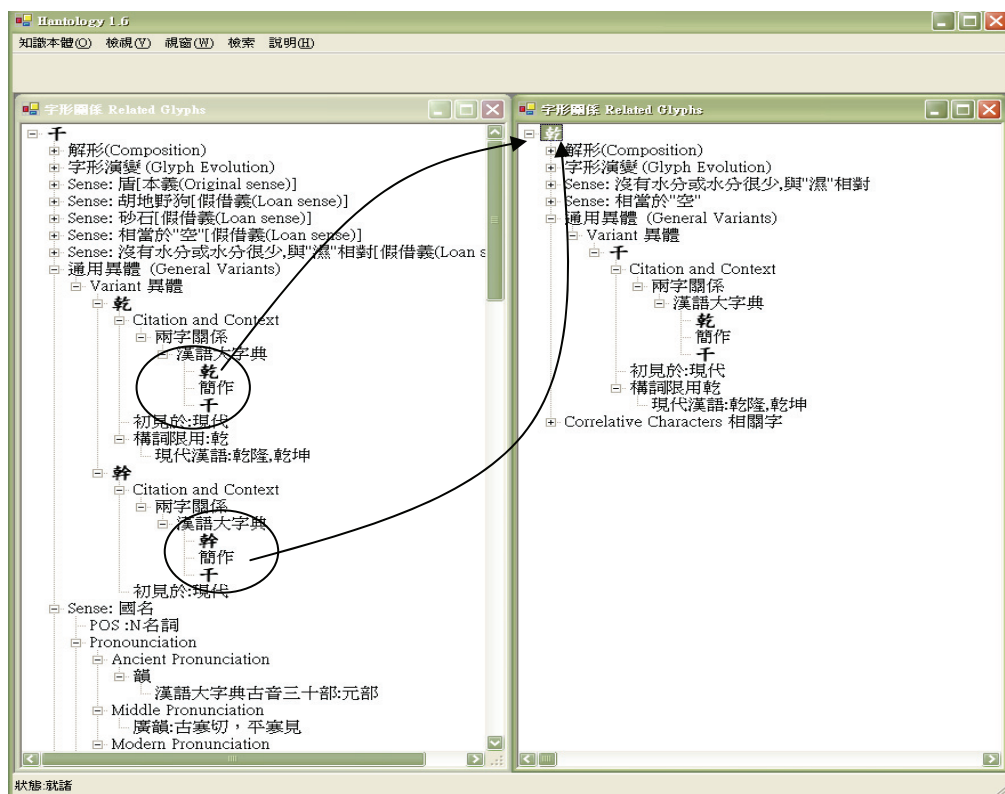
漢字知識本體可以清楚的表達正簡字之間的關係，讓華語學習者了解簡體字是否爲古今字、俗字，在什麼時候就已經與正體字有異體字關係，以及構詞的差異，這些關係在目前的漢字數位教學工具和電子字典，都沒有被適當的表達，對於學過正體後想要學習簡體字，或是學過簡體字後想要學習正體字，都能更清楚的知道正簡字之間的關係，掌握使用上的差異。



圖三、漢字知識本體描述戰國時期「干」和「乾」的通假關係

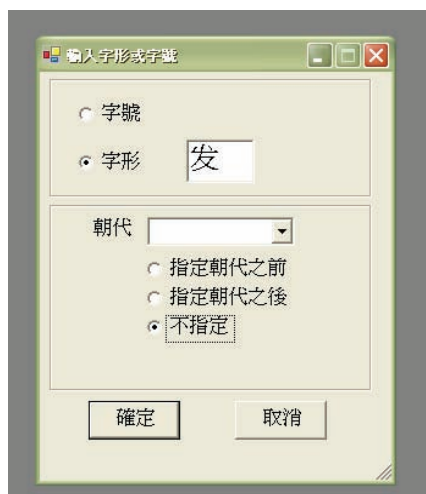
圖三以「干」和「乾」爲例，說明漢字知識本體如何描述正簡字的關係。「干」的本義是「干犯」、「冒犯」，根據中國大陸的漢字簡化方案，「乾」和「幹」都簡化爲「干」，在古漢語經常被借用爲假借字，其中就曾經被假借爲「乾」，例如《莊子》：「方將被髮而干」（裘錫圭 1995）。根據漢字知識本體

的描述，由「干」的中古音和古韻，以及「乾」的中古音和古韻，「干」與「乾」的關係為「通」，「通」在漢字知識本體中為通假關係，指的是有本字假借，「干」的字義是「沒有水分或水分很少」時，「干」為「乾」的假借字，在戰國時期有此用法，漢字知識本體並沒有它引證文獻使用「干」作「乾」，因此兩字可能在戰國有異體字的關係，後來為明確字義，不再借「干」為「乾」。直到漢字簡化方案施行，才重新建立「干」與「乾」的異體字關係，除了「乾隆」與「乾坤」的「乾」不能簡化為「干」，「乾」與「幹」都簡化為「干」。

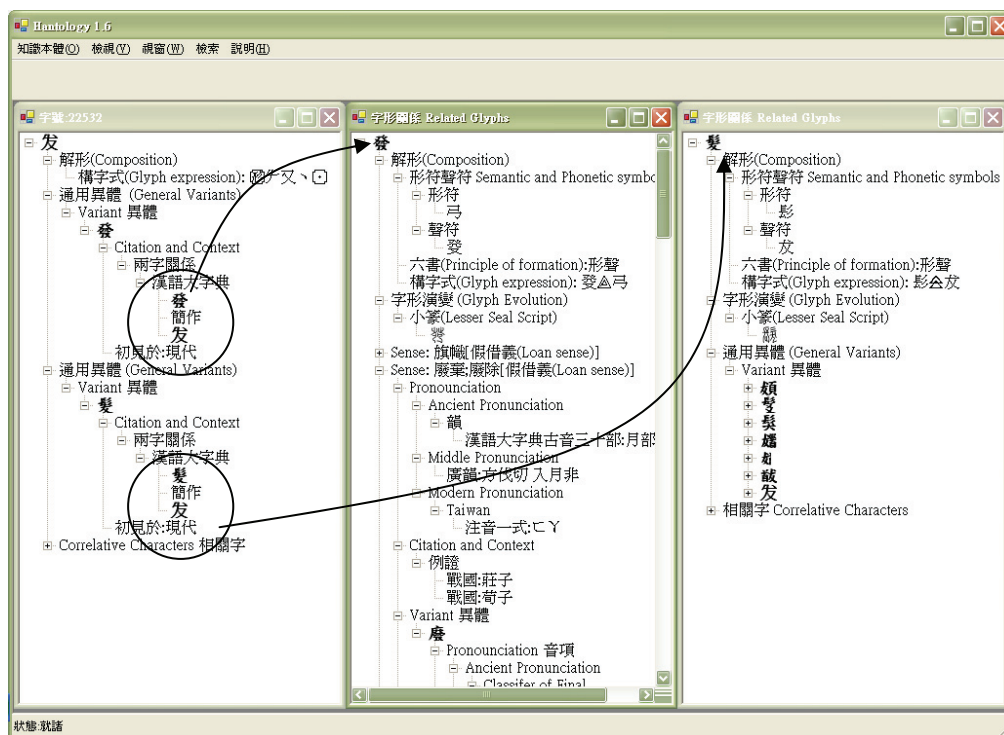


圖四、漢字知識本體描述當代「干」和「乾」的簡化關係

漢字知識本體可以利用正體字查簡體字，或使用簡體字檢索正體字，圖五是以簡體字「发」檢索漢字知識本體結果，依漢字知識本體描述，「發」和「髮」兩個正體字都是簡化為「发」。透過漢字知識本體介面分別點選「發」和「髮」，就可以呈現「發」和「髮」的知識內容。由這個例子，也同時說明了漢字知識本體可以描述正體字與簡體字之間的多對一關係。



圖五、使用簡體字「发」檢索漢字知識本體



圖六、漢字知識本體描述簡體字「发」和正體字「發」、「髮」的關係

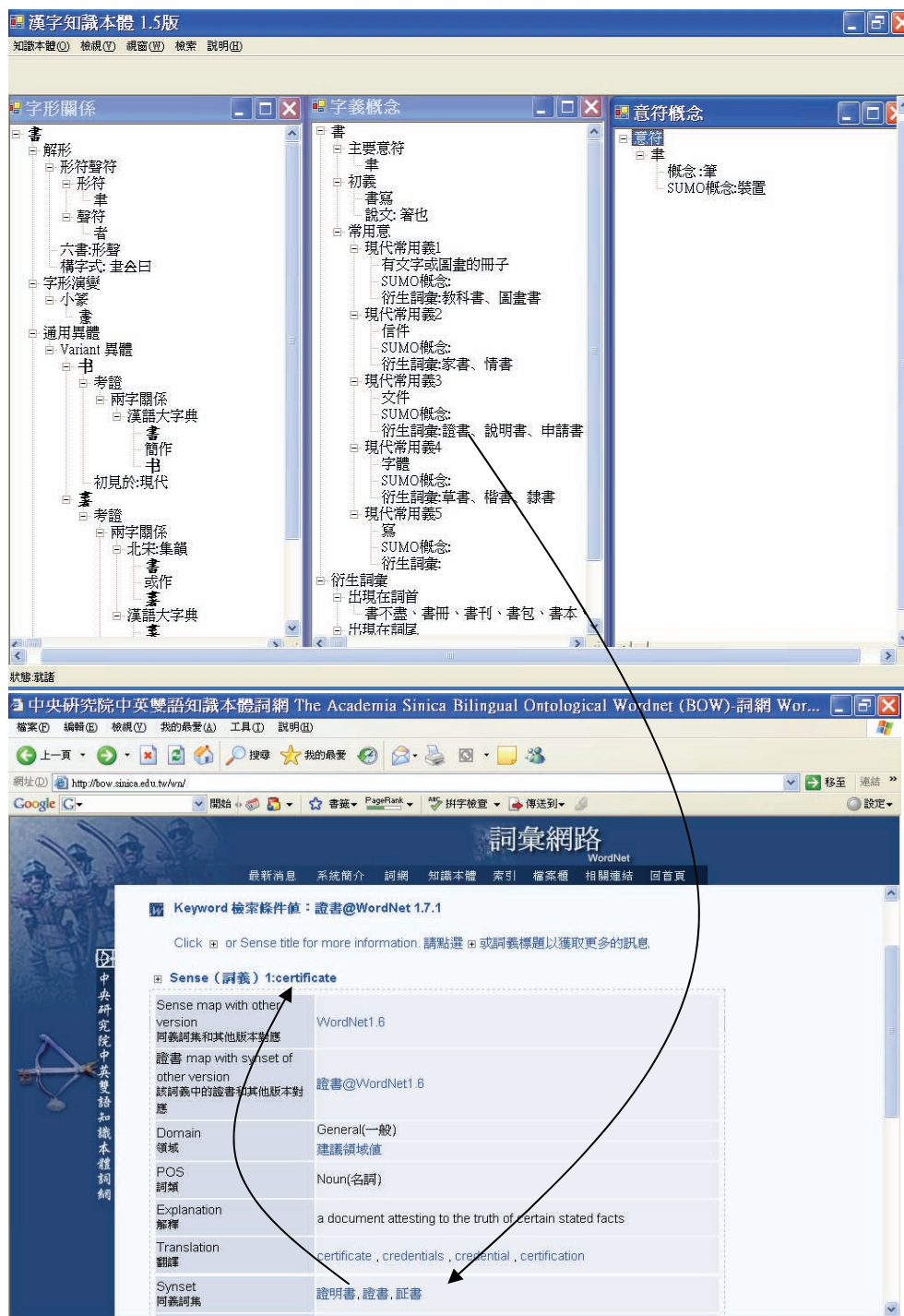
4.4 提供對比分析的學習

對比分析在語言的習得是相當重要的學習方法（黃麗儀 2007），即使在語言以外的知識學習，對比分析也是獲取知識的重要方法。漢字知識本體所表達的不只是當代的字形、字音、字義和構詞，也包含了不同時代的字形、字音、字義、異體字的變化。在語言的研究上，歷時的比較研究是很重要的方法，但是由於歷時的研究，缺乏適當的工具，要進行歷時的比較並不容易。然而對於漢字而言，在數位化時應該要表達漢字的歷時變化，才能突顯漢字的特性，幫助華語文學習者理解古籍中的內容。

利用漢字知識本體可以任意選擇兩個以上的漢字，比較其字形、字音、字義、異體字和構詞上的差異，對於華語文學習者是很方便的工具，尤其對於字義或字形相似，容易混淆的漢字，皆能夠有效地幫助華語文學習者區別差異。

漢字知識本體的每個漢字都有衍生詞，透過衍生詞與中研院中英文雙語詞網連接，經由中研院中英文雙語詞網連接到 WordNet，學習者可以學習到漢字的衍生詞，同時也可以知道這個衍生詞的英語同義詞，若英語為學習者的母語，將更能了解該漢字的使用。若是法語、德語、義大利語等為母語的學習者，則可以再經由 WordNet 連結 EuroWordNet，了解漢字詞的詞義。這些數位化語言資源的整合，可以學習和比較相同概念在不同語言的表達差異，可突破傳統的字典多語整合的困難，因為大部份字典都只提供雙語，目前華語文教學工具也很少做多語的整合。

圖七以漢字「書」為例，呈現漢字知識本體與中研院中英雙語語詞網的銜接。「書」在常用義為「文件」時，其衍生詞有「證書」、「說明書」。使用者點選「書」的衍生詞「證書」，漢字知識本體將會連接到中研院中英雙語語詞網，自動搜尋「證書」，並將「證書」的同義詞「證明書」以及 WordNet 同義的英語詞“certificate”等結果呈現給使用者，並且可以再利用中英雙語語詞網找到「證書」的上下位詞。



圖七、漢字知識本體與中研院中英雙語詞網的銜接

5. 結論

漢字數位學習工具能夠讓學習者獲得知識的途徑，主要是依賴漢字知識的數位化，雖然多媒體和網路科技的發展，讓漢字學習的方法更多元化，但是其內容深度仍待改進。長久以來，漢字數位教學工具主要以初學漢字者為對象，往往忽略已經會使用漢字的學習者，因此，所提供的漢字知識，很多都只適合初學者，缺乏較深入的內容。多數華語數位教材和線上字典，其教學對象都是初學者，但是，對於不同程度的對象，紙本的華語教材提供初級、中級和高級等不同程度的教材，數位教材也應有深淺，尤其是已經學過簡化字的學習者。相較其它的漢字數位學習工具，漢字知識本體提供內容較深的知識，不僅能夠學習到漢字的字形、字音和字義，還能夠讓華語文學習者獲得漢字的字形演變、字音的演變、字義的變化、正簡字演變和異體字關係等知識。利用漢字知識本體與中英雙語詞網的連接，可以提供漢字的學習者，由字到詞的知識，還能夠藉由漢字知識本體比較分析任何兩字的字形、字音和字義變化。

如何設計一個正體字數位教學工具，讓簡體字使用者產生學習正體字的興趣，這是重要的議題，但是缺乏廣泛的注意，原因也是多數的漢字數位教材以初學者為主，既然對象是初學者，這些工具就不會考慮使用簡體字學習者的需要。近年來，全世界學習簡體字的比例大幅增加，很多人學習漢字的優先選擇已經是簡體字，因此，漢字數位學習工具的教學對象，必需要能提供內容更深入的漢字知識，或是簡體字數位教學工具沒有提供的內容，才能吸引使用簡體字的學習者。

漢字知識本體所提供的知識，是內容較為深入的漢字知識，對於已經使用簡體字的使用者，可以直接利用簡體字檢索正體字，學習漢字意符所代表的意義，以及字形、字音和字義的演變，學習異體字之間的關係，以及學習不同朝代的使用例句。如果學習簡體字的學習者，不再只著眼於中國大陸的經濟和市場機會，開始對中華文化有興趣，因為正體字有助於學習中華文化，可能會產生學習正體字的需求，那麼正體字數位教學工具，應該要將漢字的演變與它在古漢語和中古漢語的使用情形，這些知識應融入正體字數位教學工具中，提供簡體字數位工具難以提供的教學內容，這些漢字的演變知識，很適合作為正體字教學的特色。

面對使用簡體字的人口不斷成長，如何增加正體字的使用者，不能只依賴正體字優點吸引使用者，本研究認為增加計算機處理正體中文的能力，將

是非常重要的努力方向，而增加計算機處理正體中文的關鍵，在於是否有好的數位化正體字語言資源，以提供計算機處理正體字中文資訊所需的知識。漢字知識本體的設計架構，不同於其它教學工具，它可以作為語言資源，增加在網路環境處理正體字中文資訊的能力，能夠在網際網路增加正體字優勢和學習正體字的誘因，正體字的需求增加，將吸引更多人研發優質漢字數位學習工具，提高學習正體字成效與意願，產生正向循環。

參考文獻

- Berners-Lee, T., Hendler, J. and Lassila, O.(2001) "The Semantic Web", Scientific American, Vol.284, No.5, pp.34-44.
- Chou, Y.M. and Huang, C.R.(2007) Hantology:an Ontology based on Conventionalized Conceptualization, Ontologies and Lexical Resources for Natural Language Processing, Cambridge Press.
- Huang, C.R., Chang, R.Y. and Lee, S.B.(2004) Sinica BOW(Bilingual Ontological WordNet):Integration of bilingual WordNet and SUMO. The 4th International Conference on Language Resources and Evaluation), Lisbon, Portugal, May 26-28.
- Miller, G. A.(1995) "WordNet:A Lexical Database for English", Communications of the ACM, Vol.38, No.11, Nov., pp.39-41.
- Niles,I. and Pease, A.(2003) "Linking Lexicons and Ontologies : Mapping WordNet to the SUMO Ontology", Proceedings of the IEEE International Conference on Information and Knowledge Engineering, Las Vegas, Nevada, June 23-26.
- Vossen, P.(1998) EuroWordNet : a multilingual database with lexical semantic networks, Kluwer Academic Publishers.
- 古宇時雄、前寺正彥、野村英登、穀本玲大、谷田貝常夫（1999.6），〈How 90000 Mojikyo fonts are working at present by the extension of UTF-16:文字鏡字庫與運用〉，電子古籍中的文字問題研討會。臺北：中央研究院漢籍電子文獻協調委員會。
- 石鴻珍（2006），〈多媒體網路教學漢字教學設計〉，《第八屆世界華語文教學研討會論文集》，39-40。臺北：世界華語文教育學會。

- 汪中文（2004.1），〈網路上的漢字教學與華語文學習系統〉，《漢字與全球化國學術研討會論文集》，55-63。臺北：臺北市文化局。
- 林樹（1972），《中文電腦基本用字研究》，國立交通大學工學院計算與控制學系出版。
- 周亞民、黃居仁（2006.8），〈數位典藏缺字問題的解決〉，《第五屆數位典藏技術研討會論文集》，72-77。臺北：中央研究院。
- 黃居仁、陳聖怡、周亞民（2008.1），〈語言的知識與知識的語言：由說文解字與知識本體出發的研究〉，語言研究視野的拓展國際研討會。上海：上海師範大學。
- 徐中舒（1992），《漢語大字典》。臺北：建宏出版社。
- 袁國華、曾黎明（2005），〈建立 UNICODE 漢字異體字表與異體字辭典之相關研究〉，中研院歷史語言研究所數位典藏國家型計劃（報告編號：NSC93-2422-H001-018），未出版。
- 莊德明、謝清俊（2004.1），〈漢字構形資料庫的建置與應用〉，《漢字與全球化國際學術研討會論文集》，91-102。臺北：臺北市文化局。
- 黃大一（1988），〈中文超卡：中文電腦的新領域〉，《第二屆世界華語文教學研討會論文集》，151-158。臺北：世界華文教育協進會。
- 黃沛榮（2001），〈漢學研究資料庫研發計劃－電腦古文字形〉，國家科學委員會輔助研究計劃（報告編號：NSC88-2418-H-002-025 -A9），未出版。
- 黃沛榮（2007），〈漢字教學〉，《華語文研究與教學：四分之一世紀的回顧與前瞻》，415-445。臺北：世界華文出版社。
- 黃沛榮（2005），〈漢字網路教學的字形工具〉，《全球華文網路教育研討論文集》，僑務委員會主編，354-365。臺北：世界華語文教育學會。
- 黃麗儀（2007），〈對比分析〉，《華語文研究與教學：四分之一世紀的回顧與前瞻》，329-368。臺北：世界華文出版社。
- 裘錫圭（1995），《文字學概要》（再版）。臺北，萬卷樓。
- 廖堦、曾金金（2005），〈符合 SCORM 標準的華語語音數位教學設計〉，《全球華文網路教育研討論文集》，僑務委員會主編。臺北：世界華語文教育學會。
- 舒兆民、林金錫（2007），〈多媒體華語漢字策略教學〉，第五屆全球華文網路教育研討會。臺北：僑務委員會。

- 姚道中（2007），〈多媒體與華文教學〉，《華語文研究與教學：四分之一世紀的回顧與前瞻》，489-518。臺北：世界華文出版社。
- 陳玉芳（1999），〈政府機關使用 CNS 11643 中文標準交換碼全字庫概況調查〉，《行政院主計處政府機關資訊通報》，13-17。臺北：行政院主計處。
- 陳昭容（2004），〈先秦金文簡牘詞彙資料庫〉，數位典藏國家型科技計畫－語言典藏分項計畫。中央研究院歷史語言研究所。
- 陳亮光（2007），〈多媒體與華語文教學〉，《華語文研究與教學：四分之一世紀的回顧與前瞻》，519-556。臺北：世界華文出版社。
- 楊秀芳（2003），〈漢字古今音電腦檢索系統（五）現代方音(2)〉，國家科學委員會輔助研究計劃（報告編號：NSC 90-2411-H-002-039-A9），未出版。
- 趙慧如（1991），〈漢字筆畫分類與電腦輔助教學〉，《第三屆世界華語文教學研討會論文集》，249-264 頁。臺北：世界華文教育協進會。
- 謝清俊（1992），〈談中國文字在電腦中的表達〉，《中國文字的未來》，67-82 頁。臺北：海峽交流基金會。

參考網站

- IQChinese 華語教學軟體 <http://www.iqchinese.com>
- eStroke 漢字教學軟體 <http://www.eon.com.hk/estroke>
- 五百字說華語 <http://edu.ocac.gov.tw/interact/ebook/digitalPublish2/cn500/class/shock/index2.htm>
- 漢字輕鬆學 <http://www.5qchannel.com>
- 漢字乾坤網 <http://taiwan99.tw/>
- 易學中文網路學校 <http://www.ezlearnchinese.com/chindex.html>
- 教育部常用國字標準字體筆順學習網
<http://stroke-order.learningweb.moe.edu.tw>
- 教育部國語小字典網
<http://140.111.1.43/cgi-bin/gdic/gswweb.cgi?o=ddictionary>
- 教育部異體字字典 <http://dict.variants.moe.edu.tw/>
- 中央研究院電腦漢字字形與詞彙整合知識庫
<http://chardb.iis.sinica.edu.tw/charDB3/index.htm>

[投稿：2008.11.20；審查：2009.1.8；接受：2009.03.31]

周亞民

Chou, Ya-Min

臺北市中山北路五段 250 號

250 Zhong Shan N. Rd., Sec. 5, Taipei , Taiwan

milesymchou@yahoo.com.tw

The Application of Hantology in Chinese Characters Teaching

YaMin Chou
Ming Chuan University

Abstract

Chinese characters are not only characters but often are also morphemes; therefore, they are used as the basic foundation in teaching Chinese. In computer development history, the digitization of Chinese characters is a critical advance in Chinese computers which eventually became a powerful e-learning tool to study Chinese characters.

In the past, numerous e-learning tools have been constructed to provide multiple learning approaches for Chinese characters, however, the contents provided by these tools hold various restrictions and difficulties. These problems include: insufficient knowledge on how Chinese characters evolved through its history, thus unable to link to other linguistic resources, inadequate reference among Chinese characters, oversimplify relations between traditional and simplified Chinese characters, and reckless rejection of other language forms equipped with sharing ability.

This study purposes the role of these problems in learning Chinese characters and how they can be solved by Hantology. Comparing with other e-learning tools for Chinese characters, the advantages of Hantology as a linguistic resource are its ability to improve information processing quality in Chinese characters as well as to enhance the advantage of using traditional Chinese characters in the internet.

Keywords: writing systems, Chinese language teaching, e-learning, linguistic ontology