

漢文史資料庫與斷代語言詞典編纂

徐時儀

上海師範大學

語言是人類最重要的交際工具，也是人類社會最重要的信息載體。隨著現代科學技術的高速發展，人類利用文獻信息的方式正在朝著數字化、電子化、網絡化方向發展。運用現代信息處理工具，建立大容量的漢語文史資料庫，從而對有關詞義進行多方面的統計分析，這正在成為研究漢語詞義和編纂漢語詞典的先進手段。本文擬就斷代詞典的編纂和漢語文史資料庫在斷代詞典編纂中的實用價值略作論述。

一、斷代詞典編纂的特點

語言的斷代研究和斷代語言詞典的編纂是對語言某一歷史時期的狀態進行靜態描寫。斷代的語言研究有可能對每一時段的語言進行更細更準的描寫，發現、挖掘過去未能發現或有所忽略乃至錯誤的問題，為撰寫科學、客觀、詳盡的漢語史，促進有關學科的研究打下堅實的基礎，創造必要的條件。斷代詞典是語文詞典的一種類型，即共時詞典(synchronic dictionary)。斷代詞典是對某一朝代或某一時段的語言進行歸納研究整理，¹ 然後編纂成書以供人們查檢的詞典。斷代詞典旨在真實客觀地描寫出各段語言的實況，反映出各段之間的發展演變，從而突現出語言的共時性輪廓，並在一定程度上起承上啟下的作用，為進一步編纂全面描寫詞匯演變的歷史大詞典奠定了基礎。

從理論上說，名副其實的斷代詞典只有收錄該斷代所有的詞語才能真正如實反映該時代詞匯的原貌。迄今已有的斷代詞典(如《現代漢語詞典》、《宋元語言詞典》、《上古漢語詞典》、《近代漢語詞典》等)雖有各自的斷代特色，但都沒有收錄各自所斷時代中的所

1 語言是漸變的，不一定隨著時代的更替而發生明顯的突變，因而，以朝代為依據而編纂的斷代詞典，還不是嚴格意義上按照語言分期而編纂的斷代詞典。其中有一些與語言演變的分期基本一致，有一些則儘管也或多或少反映了某些僅在該時代存在的語言現象，但往往體現不出語言演變的階段性，目前已出版的斷代詞典多以朝代為分期依據，有鑒於斷代詞典編纂的現狀，本文對此不作進一步的區分。

有詞語，因而或多或少都存在失收有關詞語的現象，有的甚至有誤收前代詞語的不足之處。顯然，斷代詞典的收詞的編纂方法僅收錄其他時代所無而為該時代獨有的詞語，而要確定哪些詞是真正屬於該時代所特有的詞，就必須先瞭解該時代詞匯的全貌，只有在對該時代主要的代表性文獻進行總帳式的詞語分析和研究之後，才能將該時代獨有的詞語挑選出來，才能避免現有斷代詞典中存在的該收未收或不該收卻誤收的通病。總帳式的編纂方法則凡是該時代使用的詞語都一概收錄，雖然沒有漏收或誤收的弊病，但往往篇幅太大，而且語言在歷史的發展演變中有許多基本詞語是歷代傳承而沿用的，這些詞語並無斷代的特色反而喧賓奪主。因而我們理想中的斷代詞典可以如同現有的斷代詞典一樣從實際功用和經濟效益著眼，只收該時代始出現的詞語和反映該時代特色的詞語，但所收錄的詞語必須在該時代使用的詞語進行窮盡性分析的基礎上來確定，即結合特色和總帳式的編纂方法，先對該時代的詞語進行總帳式的分析和研究，在瞭解和掌握該時代詞匯的全貌後，確定需要詮釋的詞語來編纂特色式的斷代詞典。

一般而言，斷代詞典的特點是按照共時性原則編纂，即對語言的某一歷史時期進行靜態描寫，收釋的詞語，從立目、定形、注音到釋義、引例，都只以某一歷史時期為限而不去涉及在此時期以前或以後的情況。詮釋對象只限於從古到今的某一階段的詞匯材料，不闡述詞形、讀音、詞義、用法等的演變。然而由於詞匯是一個開放性系統，詞匯系統實際上是匯要素及其相互關係的統一體。各斷代的詞在詞匯系統中具有一定的傳承關係，詞匯系統實際上是詞匯要素及其相互關係的統一體。各斷代的詞在詞匯系統中具有一定的傳承關係，斷代詞典對此也應有所揭示。同時，辭書的編纂畢竟還有其實用的價值，因而在辭書編纂實踐中，純粹的共時詞典是不存在的。任何共詞典都不能不酌收一些前一時期的古舊詞語，而且在許多場合還需要借助於詞源的解釋才能說明該歷史時期詞語的詞義和用法。語言的靜態描寫是研究語言發展史的出發點，描寫愈細進行語言史研究的基礎就愈扎實愈可靠。語言的靜態描寫和語史的研究是相輔相成的。靜態描寫是漢語史研究的出發點和基礎，語言史的研究又為細緻而又全面的靜態描寫提供了可能。因而斷代語言詞典的編纂必須既注意細致的動態描寫，又注意語言史的研究。斷代詞典對該時代使用的詞語的來龍去脈需要一些必要的說明，盡可能作出規律性描寫，揭示出每一個詞的詞義系統，尤其是一些常用詞語的替換。如現代漢語中「穿衣」的「穿」至遲在唐代已出現，可以用加按語的形式說明其本義是「通」，六朝時則用「著」表示，「著」的本義則為「附著」。

二、漢語語料庫的作用

編纂理想的斷代詞典，進行總帳式的詞語分析和研究是一個費時費力的浩大工程。

然而，隨著現代科學技術的發展，人類進入了數字化時代。電子計算作為現代信息處理工具，已由支持處理單一的文字或數據功能，發展至可處理聲像、色彩、動畫、圖表等多媒體功能。其容量越來越大，速度越來越高，文獻管理、檢索、統計分析、自由存取、快速傳輸等功能也越來越強大。不僅在自然科學研究領域大顯身手，並且也在社會科學研究領域裏得到廣泛的應用，進而影響到社會生活的方方面面。因而我們可以運用電子計算機來進行這項工作。

歷代的文獻記載了人類文明的進程，也反映了語言的發展和演變。我國的古籍文獻自先秦由竹簡、木牘等介質更新轉換為紙張以來，現在又面臨著文獻介質載體新的轉換，出現了一些磁盤、磁帶和光盤等電子版圖書文獻。這些電子文獻與傳統的紙質文獻相比，不僅僅是文獻載體的簡單轉換。除了佔空間小，存貯量大，檢索方便，傳輸迅速，保存期長等優勢外，電子文獻還可以利用計算機技術進行多途徑的檢索，相關信息的全方位的查閱，迅速準確的分析歸納，匯聚所需要的信息，將文本分析到字詞句等不同屬性的標注，從而建立一些電子語料庫，根據需要自動生成不同用途的數據。如利用數據庫的全文檢索功能，只要在檢索程序中輸入一個主題詞，就可以迅速地顯現其在整個數據庫中出現的次數，與其他詞語的搭配、出現的具體段落及從古到今的嬗變軌跡等情況。

我國在七十年代末期已開始著手建立漢語語料庫，迄今不少單位已研製出一系列專業、專著或專題數據檢索系統。據筆者所知，目前已研製出下列一些成果：河南大學歷史系《宋人筆記檢索系統》、《南宋主要歷史文獻全文數據庫》；四川大學古籍所《全宋文》資料檢索系統；中國社會科學院文研所《十三經》、《先秦魏晉南北朝詩》、《全上古三代秦漢三國六朝文》、《全唐詩》、《全唐文》檢索系統；蘭州大學《朱子語類輯略》引得；深圳大學《紅樓夢》全文數據庫和多功能檢索系統；哈爾濱師大《史記》全文檢索系統；山東社科院語文所《論衡》逐詞索引；陝西省中醫研究院《素問》、《靈樞》、《甲乙》、《難經》中醫典籍全文數據庫；江蘇省中醫研究所《傷寒論》、《金匱要略》、《脾胃論》等全文數據庫；山東省出版總社《二十五史》(節選)；北京大學《漢語語料庫》(五百萬字)；香港中文大學《漢及以前全部傳世文獻電腦化資料庫》和《兩漢及六朝佛經資料庫》(包括傳世典籍約一千種，共計三千三百萬字，以及竹簡帛書百餘萬字，甲骨卜辭八十餘萬字，金文十餘萬字)；武漢大學出版社《四庫全書》原格式圖文版本；上海人民出版社《四庫全書》(原格式圖文版本和可供全文檢索版本)；北京國學時代文化傳播有限公司《國學寶典》；廣西大學《古今圖書集成索引》；北京超星電子技術公司《古今圖書集成》。上海圖書館有古籍善本光盤，已收錄館藏善本《社工部集》等三千多種，可供上網檢查查閱。台灣中央研究院歷史語言研究所從1984年開始推動「史籍自動化計劃」，首先開發了《二十五史全文資料

庫》，繼而又逐步建立了先秦兩漢史籍、《十三經注疏》、《諸子》等大型資料庫，已可在因特網上查詢使用。1991年，國家語言文字工作委員會開始建立國家級的大型漢語語料庫，以推進漢語的詞法、句法、語義和語用的研究，同時也為中文信息處理的研究提供語言資源，計劃其規模將達七千萬漢字。全部語料分人文與社會科學類(佔59.6%)、自然科學類(佔17.24%)、綜合類(佔9.36%)、報紙等類(佔13.79%)。中國基本古籍庫光盤工程已取得重大進展，選有10000多種文獻，涵蓋了全部中國歷史和文化，其內容相當於3部《四庫全書》。² 日本花園大學國際禪學研究所也於1995年出版了電子達摩，收入《臨濟錄》、《宛陵錄》、《六祖壇經》、《馬祖語錄》、《古尊宿語錄》、《五燈會元》等七十多種禪宗語錄。中華電子佛典協會2002年出版了CEBTA電子佛典系列，收錄有《大正新修大藏經》第一冊至第五十五冊暨第八十五冊。

江藍生先生《近代漢語斷代語言詞典系列》序說：「要編寫一部斷代語言詞典，首先要對該時代的文獻(主要是反映口語成分較多的白話資料)做比較全面系統的調查，在掌握大量第一手可靠資料的基礎上，進行選擇、整理和考釋，而要做到這一點又談何容易！」³ 誠如江先生所說，研究考釋詞語和編纂詞典，往往有三分之二甚至更多的時間是花在語料的搜集、整理、排比、分析、歸納的工作上，而現在計算機則可根據我們編制設定的程序迅速準確地完成這些工作。如果說電子語料庫的使用相當於傳統研究工作中詞語卡片的抄錄，那麼一些自動分詞、自動斷句、音韻分析、語義分析、語法分析的專用軟件的就相當於詞語卡片的抄錄，那麼一些自動分詞、自動斷句、音韻分析、語義分析、語法分析的專用軟件的就相當於詞語卡片的排比、分析、甄選、歸納和統計。1970年美國人工智能專家W. A 伍茲提出一種自動處理自然語言的方法，稱為「擴充轉移網絡」，簡稱ATN。用擴充轉移網絡來分析句子，解釋語義，便可求得對自然語言的理解。目前也有一些基於自然語言處理漢語的軟件。1986年北京文獻服務處(BDS)研製了基於自然語言處理的中文情報檢索和處理系統CIRPON。1990年北京信息工程學院與《人民日報》社合作開發了全文檢索系統BITIFIRS(full text retrieval system)。電子部計算機與微電子技術發展研究中心(CCID)中文信息處理開放實驗室(CIOOL)張潮生等研製了中文全文檢索系統TIR，可對各種文本型資料和某些數據的文件進行操作。湖北大學正在開發研製「《詩經》語言分析全息系統」將可以對《詩經》的語音、詞匯、語法、修辭等各方面進行全方位的分析研究。有了這樣一些專用軟件和數據庫就能使詞語的研究由籠統的定性分析走向科學的定量分析。如「吹」有風吹義和吹樂義，根據對《詩經》、《莊子》、《論衡》、《世說新語》和李白、杜甫等唐詩語料的統計分析，可以看到先秦時「吹」的吹樂義佔多

2 張艷：《萬種古籍進光盤》，《文匯報》2001年3月19日第6版。

3 江藍生：《近代漢語斷代語言詞典系列序》，江藍生和曹廣順：《唐五代語言詞典》(上海：上海教育出版社，1997年)。

數，可是到了唐代李白和杜甫的詩歌中，「吹」的風吹義不僅佔多數，而且幾乎是吹樂義的兩倍。⁴ 以《詩經》、《莊子》、《論衡》、《世說新語》和唐詩等語料庫的分析為依據，編纂斷代詞典時對「吹」這兩個義位的變化就可描寫得更科學更客觀。

三、利用數據庫編纂斷代詞典的一些設想

計算機在處理海量的語料方面具有超人的優勢，但計算機並不能取代人的所有工作。斷代詞典的編纂更重要的工作是對語料年代的判斷和詞義的分析詮釋。利用數據庫編纂斷代詞典實際上是一種人機互動的過程。計算機根據我們設定的條件對大量語料進行處理，只要符合條件，就會運程序給出結果，必然和偶然，規律和例外，通例和特例，皆不加區別，一視同仁。這就需要我們對運算所得的結果進行分析，發現問題，進一步設定條件，讓計算機對語料進行再處理，剔除特例和例外。這樣不斷地發現問題，設定相應的程序讓計算機按照要求解決問題，通過人機互動，反覆運作，將我們對語料的分析判斷與計算機對語料的形式化處理結合起來，最終就能得到預期的結果，編纂成在窮盡某時期所有文獻詞語的總帳式分析基礎上對具有該時期特色的詞語進行詮釋的理想斷代詞典。利用數據庫編纂斷代詞典大致可分如下四個步驟：

1、確定斷代分期

確定斷代，即根據語言發展的內在規律和演變的階段性，兼顧歷史朝代的興亡更替，從漢語發展的歷史長河中橫切一刀，劃分出斷代的時間範圍。如可分上古、中古、近代三個時段，也可分為先秦、秦漢、六朝、唐五代、宋代、元代、明清七個時段。⁵

2、選擇典型語料

全面地瞭解各個時段的文獻概貌，兼容並包不同的文體和內容，選擇一些具有代表性的文獻，建立每一部文獻的專書語料庫。所謂代表性主要看其是否使用或反映了該斷代的口語，因為有書面記載以來的整個漢語史都是口語和文言並存的歷史，只有記載口語的文獻才比較真實地反映了斷代所用詞語的面貌。如漢代有《史記》、《論衡》、樂府民歌和漢譯佛經等代表性文獻，唐五代有敦煌變文、曲子詞、禪宗語錄、《祖堂集》及李白、杜甫、白居易、寒山、王梵志等的詩作等代表性文獻。

斷代詞典旨在真實客觀地描寫各段語言的實況，因而選擇各個時段中具有代表性的專著時需要作一番辨偽工作，確定所選專著屬於該斷代，注意避免「時代的錯誤」，即錯

4 楊克定：《李白詩歌中的自然V》，程湘清：《隋唐五代漢語研究》（濟南：山東教育出版社，1990年），頁390。

5 關於漢語史的分期問題，諸家的說法不一。蔣宗許和李潤《試論斷代語言詞典的編纂》一文探討了斷代語言詞典編纂的斷代、收詞和釋義可參看，本文從略。蔣宗許和李潤：《辭書研究》，第4期（1991年）。

誤地把後一時期的語言現象當作前一時期的語言現象。如入矢義高先生在《中國口語史的構想》一文中說，「遮莫」這個詞在詩中常常使用，可不知為甚麼在散文中幾乎不使用。它是一種詩的詞匯，意思是「儘管」，以往日本人讀作「本非如此，只能如此」之意，其實這是不恰當的。這是一個很奇妙的詞，從字面上看也無法理解會有這種意思。「遮莫」這個詞，從唐代一直到清朝為止，一直是固定地作為詩的用詞，然而最早出現的文獻是晉代的《搜神記》，此後一直到唐代初期的《遊仙窟》中再出現，這中間的一段歷史是空白的。到底是使用過，還是雖然使用過而在文獻裏沒有被記載，這一點不明確。也許就是這個原因，宋代人好像已經不理解這個詞，在注釋唐詩時還對這個詞專門注解有各種各樣的說明，一般把它解釋為「儘教」。⁶ 其實《搜神記》原書傳至宋代就已散佚，今傳本是明代胡元瑞從《法苑珠林》及諸類書中輯錄而成。江藍生先生《八卷本搜神記語言的時代》一文指出「遮莫」最早見於7世紀末8世紀初的文獻中，在魏晉南北朝的資料中還沒有出現，其基本意義是表示縱予，既可作副詞，也可作連詞，這是唐代開始習用的口語詞。⁷ 從版本上可證《搜神記》中的「遮莫」一詞實際上是後人闖入的後時資料，不足以成為晉代已有此詞的依據。入矢義高先生指出此詞從晉代的《搜神記》一直到唐代初期的《遊仙窟》這中間的一段空白也反映了此詞唐以前尚未見使用。因而《搜神記》一書雖題為晉代干寶所作，但不能用作反映晉代的語言材料，又如現存唐代禪宗語錄大多是後世編纂的，難免會有後人的改動之處，有的甚至經過多次改訂。如《壇經》有中唐的敦煌本、宋初的惠昕本、北宋的契嵩本、元代的宗寶本和明藏本等不同版本，其間差異頗大，後出的本子對前面的本子有改，有刪，有增。現存最古的敦煌寫本《壇經》已非原來面貌，大約已是五代寫本，後出的本子與敦煌寫本相較而言，改動當然也就更多了。如明藏本《六祖壇經·機緣第七》中有「恁麼」、「甚麼」等，然而《壇經》的敦煌本，覆宋本卻沒有這些詞，我們就不能將這些詞收入唐代語言詞典中。因而，編纂斷代詞典時在選擇具有代表性的文獻時有必要對所選語料先作版本的考證和文本的校勘，避免使用已闖入後時資料的文獻。

文獻的整理首先要比較所選定的代表性文獻的各種版本，確定作為收詞依據的原本，然後將校勘底本輸進電腦，以其他版本參照考訂、校點。利用文字插入和移動功能，邊讀邊校。文字校勘與分詞斷句的得當與否，直接關係到數據庫質量的高低。如《世說新語》一書有各種版本，文字差異頗大，「是正文字，勘定文句」的校勘工作是必不可少的。⁸ 又如《全唐詩》成書於清康熙年間，由於成於眾手，加之所據版本紛繁，其中異文、訛誤頗多，校勘工作自然成為建立《全唐詩》數據庫的首要任務。每校勘完一部文獻，即

6 入矢義高：《中國口語史的構想》，《集刊東洋學》第五六號（1986年），後收於《空花集——入矢義高短篇集》（1992）。

7 江藍生：《八卷本搜神記語言的時代》，《中國語文》第4期（1987年）。

8 蔣宗許：《世說新語詞典編纂瑣談》，《辭書研究》第3期（1991年）。

建立一個相應的語料庫，並編制分析詞語的有關程序，最終形成一個反映某個時代詞語資料長編的數據庫。

3、進行數據分析

斷代詞典應該對該時代所有的詞語進行全面研究，有了反映某個時代詞語資料長編的數據庫，就可利用電子計算機所具有的逐詞、逐句等分類排列組合的功能，對數量浩繁、散見各代的文獻資料進行總帳式的數據統計，綜合比較某些詞語在各代的使用頻率或用法上的區別。有些音義相同而用字不同的詞語，如元曲中「艷」、「焰」又寫作「釅」，則可根據數據庫的分析從形音義角度說明「艷」、「焰」是「釅」的同音假借字。

「一部斷代的語言詞典不同於一般的語文詞典、百科詞典，在有限的篇幅裏要盡量收錄那個時代的新詞新語和特有詞語，同時又要能得其輪廓，反映該時代的詞匯系統和整個面貌。」⁹ 每個時代都有其代表性的詞匯，我們可以採用窮盡性的選擇性相結合的原則來分析某個時代詞語資料長編數據庫中的詞語。斷代詞典要在對數據庫中所有詞語進行窮盡性分析和研究的基礎上從中篩選確定具有該時代特色的詞語，包括某個斷代的所有詞匯和某一詞語在該斷代的不同義項，如「修理」在唐代有「處置，料理」義，敦煌變文《舜子變》：「緣人命字重，如何但修理他？」又有「整治飲食」義，《西陽雜俎》卷七稱，有一將軍家，「善均五味，嘗取敗障泥胡祿修理食之，其味極佳。」

我們可根據某些詞語的數據分析來確定新詞新義出現的始見書和首見例。如石毓智和李納先生《十五世紀前後的句法變化與現代漢語否定標記系統的形成——否定標記「沒(有)」產生的句法背景及其語法化過程》一文說：「『沒』在唐、宋時期一直是一個名詞性成份的否定標記，它否定動詞性成份的用法十五世紀前後才出現。」¹⁰ 「明中葉(約十五世紀)以後，『沒』開始逐漸用於謂語中心動詞的否定。」¹¹ 「『沒』從約八世紀起開始引伸作否定『領有』的動詞，一直到約十四世紀的五、六百年的時間 都是單獨用，『沒有』是後起的用法。這跟現代漢語中『沒』和『沒有』可以自由替換的情況不同。根據梅祖麟(1984)對宋末元初大量文獻調查，那時還是只用『沒』不用『沒有』，編於十四世紀到十五世紀的《老乞大》、《朴通寧》、《水滸傳》才開始用『沒有』。所以『沒有』的出現應大致在十四世紀左右。這是一個重要的發展，因為這表明『沒』由一個動詞變成動詞『有』的單純否定標記，這為『沒』向否定動詞性成份的擴展提供了可能。」¹¹ 「『沒』作為普通動詞的否定標記的時間比『沒有』凝固成複合動詞的時間大約要晚一、兩百年的時間。」¹¹ 「十六世紀以後現代

9 同注3。

10 太田辰夫：《中國語歷史文法》(北京：北京大學出版社，1987年)認為「大約始於元明」，孫錫信：《漢語歷史語法要略》(上海：復旦大學出版社，1992年)也認為「用於否定動詞大約也在元明之際」。

11 《漢語大詞典》釋「沒(mò)」的副詞義為「表示否定，相當於莫、不要」。引明吾丘瑞《運甓記，諸賢渡江》為最早書證。「沒(méi)」的副詞義為「未，不曾」。引《紅樓夢》為最早書證。

漢語中的作為名詞性和動詞性否定標記的『沒(沒有)』已經形成了。」¹² 石毓智和李納先生所根據的梅祖麟先生對宋末元初大量文獻調查出自梅先生《從語言史看幾本元雜劇賓白的寫作時期》一文，梅先生在文中指出：「現代漢語的『沒有』有兩個用法，一個相當於文言的『未』，「一相當於文言的『無』，他根據高歌蒂先生《朱子全書內所見之宋代口語》一文而認為《朱子語錄》根本沒有「沒有」這個語詞。¹³ 高歌蒂撰《朱子全書內所見之宋代口語》時尚不能用計算機來進行語料的統計，實際上《朱子語錄》卷二十一《論語三·道千乘之國章》所載葉賀孫記錄的朱熹解說《論語》時已說到：「『道千乘之國，五者相因，這只消從上順說。人須是事事敬，方會信。才信，便當定如此，若恁地慢忽，便沒有成。』」¹⁴ 《朱子語類》一書的數據分析為編纂斷代詞典提供了「沒有」一詞的始見書和首見例。

我們還可根據對各代所使用詞語的分析來確定某些詞語或詞義不為後代沿用的消失年代。如「思」讀作平聲，義為思考、考慮，《廣韻》：「思，思念也。息茲切。」引伸指懷念、想念。讀作去聲則有三個意思。一為考慮問題周密深入而又敏捷。《尚書·堯典》：「昔在帝堯，聰明文思，光宅天下。」陸德明《經典釋文》：「思，息嗣反，又如字。」二為對外界所作的考慮或考慮結果、想法等。《詩經·唐風·蟋蟀》序：「本其風俗，憂深思遠。」三為因感於外界而內心哀愁、愁怨。《毛詩序》：「治世之音安以樂，其政和；亂世之音怨以怒，其政乖；亡國之音哀以思，其民困。」「思」的「哀愁」在中近古尚沿用。如白居易《琵琶行》：「弦弦掩抑聲聲思，似訴平生不得志。」又如李商隱《錦瑟》：「錦瑟無端五十弦，一弦一柱思華年。」金人元好問《論詩雜句》中說：「望帝春心托杜鵑，佳人錦瑟怨華年。」元好問改「思」為「怨」透露出其時「思」的「哀愁」義已罕用，這是進行斷代語料數據分析時值得注意的語言現象。斷代詞典的編纂只有從歷時著眼，從共時著手，堅持窮盡性與選擇性相結合的原則，才能充分地體現出斷代詞典所具有的時代特色。

4、描述斷代詞義

斷代詞典最主要的工作就是對斷代所使用的詞語進行共時的描寫。語言既有創新性，又有繼承性。許多詞語在幾個時代都通用，只有對大量語料進行比較分析，才能發現這些詞語之間的細微區別，只有對所有的詞語進行具體全面的描寫，才可能反映出斷代語言的特色。王力先生在《新訓詁學》一文中說：「從歷史上觀察語義的變遷，我們首先

12 石毓智、李納：《十五世紀前後的句法變化與現代漢語否定標記系統的形成——否定標記「沒(有)」產生的句法背景及其語法化過程》，《語言研究》第2期(2000年)。

13 梅祖麟：《從語言史看幾本元雜劇賓白的寫作時期》，《語言學論叢》第十三輯，北京：商務印書館1984年，頁133。

14 黎靖德：《朱子語類》(北京：中華書局，1986年)，頁496。

應該有明銳的眼光，任何細微的變化都不能忽略過去。」¹⁵ 只有通過對該斷代中所有詞語及其運用狀況分門別類的描寫、分析，才能揭示出共時平面上這些詞語和語言現象的內在聯繫和語言發展的規律。如漢魏六朝的文人詩中常見「雲雨」一詞是由比喻方式構成的新詞，義為「分別、別離」，實際上是「雲飛雨散」等詞語的簡縮，而由宋玉《高唐賦序》中神女「旦為朝雲，暮為行雨」的句子簡化成的「雲雨」則可指對佳人的思念，也可代指所依戀的女子。大約到了宋元時期，由《高唐賦序》中的「雲雨」附會而有「男女歡會」義。¹⁶

斷代詞典的編纂一方面注重靜態的描寫，另一方面也要注意進行歷史的動態的縱向比較，說明詞語的來龍去脈，在縱向和橫向的綜合比較中注意對常用詞語進行多向性全方位的考察。語言本身是一個系統，「語言的歷史發展也是系統的。從一個時代到另一個時代，是一個新的系統代替一個舊的系統。它不是零零碎碎地變的，所以我們研究語言史決不能零敲碎打，而必須對整個語言系統進行全面的審查。」¹⁷ 如表示「飲食」義的動詞，上古用「食」、「飯」、「飲」、「餐」等；中古出現了「吃」，但「食」、「飯」、「飲」、「餐」等仍在使用。《世說新語》中「食」和「飲」用得較多，「吃」、「飯」、「餐」偶見，李白詩中「食」和「餐」各20例，「飯」5例，沒有用「吃」。白居易詩中「食」70多例，「餐」17例，「吃」6例，「飯」2例。¹⁸ 圓仁《入唐求法巡禮行記》中「食」3例，「飯」15例，「吃」64例。¹⁹ 然而質變都表現為一定的數量，編纂斷代詞典時可用加按語的形式從史的角度對這些詞語的量變（語言現象出現的頻率）和質變（具有斷代意義的語言差別）過程加以描寫說明，在某種程度上反映各個時段詞語之間的內在聯繫。

四、結語

斷代語言研究和斷代詞典編纂是漢語史研究的基礎，而多功能的大型漢語詞義語料庫的建立又是斷代語言研究和斷代詞典編纂的基礎。在斷代詞典的編纂中，文史資料庫的利用可以大大提高編纂工作的效率，使編纂者從繁重的摘錄、排比卡片的手工勞動中解放出來，集中精力對語言現象和其本質進行研究。歷代流轉下來的文獻浩如煙海，如果不利用計算機逐部建立文史資料數據庫，再逐詞綜合比較分析，那麼編成的斷代詞典難免掛一漏萬。因此，我們面臨的任務是系統而有選擇地先將一些重要的、有代表性的文獻輸入計算機，建立有關這些文史資料的語料庫，進而制定編纂系列斷代詞典的規

15 王力：《新訓詁學》，《開明書店二十周年紀念文集》，1947年，又載《龍蟲並雕齋文集》第1冊（北京：中華書局，1980年）。

16 王云路：《「雲雨」漫筆》，《古漢語研究》第3期（2000年）。

17 王力：《我的治學經驗》，《語言學論文集》（北京：商務印書館，1985年），頁10。

18 蔣紹愚：《白居易詩中與「口」有關的動詞》，《語言研究》第2期（1993年）。

19 董志翹：《入唐求法巡禮行記詞匯研究》（北京：中國社會科學出版社，2000年），頁231。

劃，形成一個既反映各個時代漢語的面貌特點，又彼此相輔相成反映各代詞語發展演變脈絡的語言系統，揭示出漢古今演變的脈絡，從而推動漢語史研究的深入進展，並在此基礎上編纂歷史性大詞典，進而也為基於漢語的智能化計算的設計奠定了堅實基礎，為佔世界人口五分之一的漢語使用者進入信息化時代開關通途達徑。

我國自編纂第一部詞典《爾雅》起，一直是採用人腦加手工的方式，辭書編纂者終日埋頭於典籍和詞語卡片中，十年甚至數十年才磨一劍。現在，信息技術的飛速發展為辭書編纂和漢語史的研究提供了以往所沒有的機遇，詞義的分析和辭書的釋義可以通過建立數據庫點擊鼠標米進行，文史資料庫建設正在成為辭書編纂質量的先決條件和重要基礎。如果說上世紀末利用計算機編纂辭書還使人覺得神奇的話，那麼當新世紀的曙光普照寰球之際，這已成為辭書編纂的必要技術手段。人類已經步入一個數字化和網絡化的時代，新時代的辭書編纂前景最引人入勝的無疑將是古老傳統文獻與當今計算機技術的珠聯璧合，新時代的系列斷代詞典編纂無疑也將充份地體現辭書編纂中古與今的完美、和諧、巧妙的融合。

【本文屬專著類】