

# 基於漢語歷時語料庫的雙音同素逆序現象研究\*

熊丹<sup>1</sup>, 陸勤<sup>1</sup>, 羅鳳珠<sup>2</sup>, 石定栩<sup>3</sup>, 趙天成<sup>1</sup>

香港理工大學<sup>1</sup> 電子計算學系<sup>3</sup> 中文及雙語學系<sup>2</sup> 台灣元智大學中國語文學系

<sup>1</sup>{csdxiong, csluqin, cstschiu}@comp.polyu.edu.hk <sup>3</sup>dingxu.shi@polyu.edu.hk <sup>2</sup>gefjulo@saturn.yzu.edu.tw

**摘要:** 鑒於同素逆序的構詞現象在漢語詞彙發展中的重要性及其對計算機自動分詞的影響, 本工作從自然語言處理的角度對歷時語料庫中的雙音同素逆序詞進行比較研究, 該語料庫包含兩部已完成分詞和命名實體標註的先秦典籍和四部明清小說。本文根據對歷時語料庫中雙音同素逆序詞的窮盡統計, 針對史籍和小說這兩種不同體裁文本中的使用情況進行比較分析, 其結果有助於考察這些詞在不同體裁文本中的使用傾向, 及在不同時代的沿襲和演變。

**關鍵詞:** 雙音同素逆序詞, 漢語歷時語料庫, 中文分詞

## Morpheme Inversion in Disyllabic Compounds — Cases in Chinese Diachronic Corpora

Dan Xiong<sup>1</sup>, Qin Lu<sup>1</sup>, Fengju Lo<sup>2</sup>, Dingxu Shi<sup>3</sup>, Tin-shing Chiu<sup>1</sup>

<sup>1</sup>Department of Computing, <sup>3</sup>Department of Chinese & Bilingual Studies, The Hong Kong Polytechnic University

<sup>2</sup>Department of Chinese Linguistics & Literature, Yuan Ze University, Taiwan

<sup>1</sup>{csdxiong, csluqin, cstschiu}@comp.polyu.edu.hk <sup>3</sup>dingxu.shi@polyu.edu.hk <sup>2</sup>gefjulo@saturn.yzu.edu.tw

**Abstract:** Morpheme inversion is a significant lexical phenomenon in the historical development of Chinese words, and it poses additional difficulties in Chinese segmentation, especially in the processing of Chinese classics. This work conducts a study on the disyllabic morpheme-inverted compounds in the Chinese diachronic corpora from the perspective of natural language processing. The corpora include two pre-Qin classics and four notable novels created in the Ming and Qing dynasties, in which words are segmented and named entities are tagged. Based on the full statistics and analysis, a comparative study is carried out on the use of the disyllabic morpheme-inverted compounds in the two styles of Chinese texts, that is, historical works and fictions. The results demonstrate the preference of the inverted words used in the two types of texts as well as the evolution of these compounds in Chinese classics.

**Keywords:** Disyllabic morpheme-inverted compounds, Chinese diachronic corpora, Chinese segmentation

### 1 前言

無論在古代漢語還是現代漢語中, 詞語的同素逆序都是非常顯著的現象, 在語言學研究和漢語教學中也頗受重視。在使用計算機進行自動分詞時, 同素逆序詞會造成一定的干擾。本工作從自然語言處理的角度出發, 基於「歷代語言知識庫建置計畫」所建置的語料庫[1], 對其中的雙音節同素逆序詞進行分析。該語料庫包括已完成分詞和命名實體標註的《左傳》、《國語》兩部先秦典籍及《三國演義》、《水滸傳》、《金瓶梅》、《紅樓夢》四部明清小說。通過對該語料庫進行窮盡統計, 觀察不同時代、不同體裁文本中同素逆序詞的分布概貌, 並對先秦典籍和明清小說這兩種不同體裁文本進行比較分析, 找到了一些極具參考價值的結果, 可以幫助我們考察這些詞在不同體裁文本中的使用傾向, 以及在不同時代的沿襲和演變。本文的第 2 節介紹同素逆序詞的研究概況, 第 3 節介紹研究範圍, 第 4 節展示雙音同素逆序詞在語料庫中的分布, 第 5 節基於數據和實例對兩種不同體裁中雙音同素逆序詞的使用情況進行比較分析。本文所列舉的實例全部來自「歷代語言知識庫建置計畫」所建置的語料庫(下文稱為「歷時語料庫」)。

### 2 研究概況

中文同素逆序詞是語言學界廣泛關注的現象。現有的研究成果頗為豐碩, 主要側重在共時研究。研究的對象包括單部專書, 詞典, 某位作者的系列著作, 某一對或某一類特定的同素逆序詞, 以及不同文體類別的文本等; 研究的角度多種多樣, 包括詞義辨析、詞彙探源、修辭、方言、跨

\* 本文承 2010 年蔣經國國際學術交流基金會「歷代語言知識庫建置計畫」(RG013-D-09) 部分資助。

語言比較、跨文化（涉及中、韓、日）比較、漢語教學、辭書編纂等，參考文獻[2]、[3]以及[4]的第 6-9 頁對此有比較詳盡的綜述。總體看來，從歷時的角度進行的研究相對比較薄弱，典型的代表作包括[4-6]。雖然有些是明清小說中同素逆序詞的研究，但都以針對單本專書為主。本工作從自然語言處理的角度出發，基於前述已完成分詞和命名實體標註的六種文本，通過對歷時語料庫中雙音節同素逆序詞的窮盡統計進行對比研究，因此研究對象和側重點均與已有研究不同。

### 3 研究範圍

本文的研究範圍為歷時語料庫中的同素逆序詞，但不包括明清小說中的韻文。這是因為詩詞的體裁較為特殊，出現詞素倒序常常是要創造新詞，以遷就某一特殊的格律要求，或者是要達到某種修辭效果，此類情況宜單獨研究。

本文所研究的雙音節同素逆序詞是指構詞語素相同，但語素序位互逆的雙音節合成詞。語言學界對於同素逆序詞的界定方法有寬有嚴，並沒有統一的定義，而且也沒有統一的名稱，常見的叫法還有「顛倒詞」、「同素反序詞」、「同素異序詞」等等。本文參考主流看法（見參考文獻[4,7-9]），針對我們的研究目的和歷時語料庫中詞彙的特徵確定了自己的標準。下文首先從「逆序」和「詞」這兩方面初步明確我們的研究對象，然後通過實例來界定「同素」。對字形、語音、語素義和詞義的考證參考了一些詞典[10-13]，及兩部先秦典籍和四部明清小說中詞彙的現有研究成果，包括辭書、專著和學位論文[14-24]。

#### 3.1 關於「逆序」和「詞」

「逆序」是指從詞的外形特徵看，其兩個構詞語素順序互逆。由於本文的研究對象是雙音節詞，只包含兩個構詞語素，因此不涉及學術界關於多音節詞是全部語素顛倒或部分語素顛倒的爭議。

對於「詞」的界定，本工作的標準和現有的研究相同的是排除由一個語素構成的雙音節單純詞。這些詞雖然有兩個音節，但單個音節並沒有意義，所以在分詞時不會造成困難，例如：

唧咕—咕唧                  蹊蹺—蹺蹊                  靦腆—腆靦

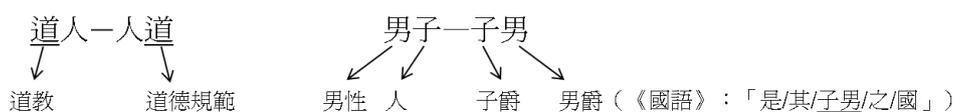
由於本工作是從自然語言處理的角度出發，在建構歷時語料庫時，我們的分詞原則是語義為主要原則，以語法為輔助原則，基本原則為「致力於在做到切分後不造成語義丟失、轉換、引申或歧義的情況下，切分到最小完整語義單位」[1]。該分詞規範基本沿用了北京大學的現代漢語分詞體系[25]、並借鑒了台灣中央研究院的分詞標準[26]。基於這些規範，結構固定、意義固化的詞組、習語、成語等不予切分，作為一個分詞單位，即一個「具有確定的語義和語法功能的基本單位」[27]。正因為如此，語料庫中的詞條不僅包括語言學意義上的「詞」，還包括部分詞組，例如「心多」是主謂結構的詞組，詞義為疑心重，如《紅樓夢》中的例子「況/兼/黛玉/心多」。由於切分為「心/多」之後易造成語義丟失，我們將其作為一個分詞單位，所以「心多—多心」也在我們的研究範圍之內。盡管我們沒有在嚴格意義上給「詞」下定義，但為了方便描述，在本文中還是將這些結構統一稱為「同素逆序詞」。

#### 3.2 關於「同素」

「同素」要求兩對構詞語素同義、同形、同音。同義可以是語素義完全相同，也可以是相近或具有明顯關聯，例如：

驛館—館驛                  糧米—米糧                  較計—計較

根據這一原則，語素義無明顯關聯的疑似「同形逆序詞」就排除在外，因為這些詞的其中一對語素義沒有明顯關聯，有些甚至兩對均無關聯，例如：



有些構詞語素同音同義但不同形，主要包括異體字、通假字、古今字，例如：

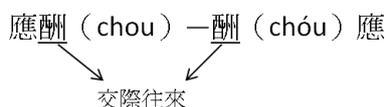
熟嫻—嫻熟                  報荅—荅報

這些詞理應歸入「同素逆序詞」之列，但建立識別異體字、通假字、古今字相應的知識庫需要整合，未來建立完成之後，將再加入這一類詞彙的分析。

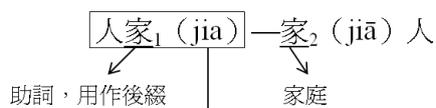
同形而不同音的字，通常會表示不同的語素義，所以一般不能作為同素逆序詞，例如：

家當 (dàng) — 當 (dāng) 家                      好 (hào) 學 — 學好 (hǎo)

輕聲是一種特殊的語音現象。有些語素通常情況下總是發輕聲，如作為動態助詞的「了」，作為結構助詞的「地」、「得」、「的」等。有些語素並非總是發輕聲，而是只在某些特定的複合詞中發輕聲，而且有些輕聲語素的意義與非輕聲語素不同。如果輕聲不造成語素義的變動，則不作為判定同素逆序詞的考慮因素。例如「應酬」中的「酬」通常發輕聲，但在台灣國語中不發輕聲<sup>1</sup>，輕聲與否並不改變「應酬」的基本意義，所以下面一組是同素逆序詞。

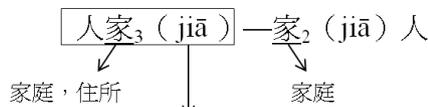


如果輕聲造成了語素義的區別，發輕聲和不發輕聲的就不是同一個語素。例如下面一組不是同素逆序詞：



人稱代詞，指自己，別人，某個人或某些人，如《紅樓夢》中的例子：  
1) 「人家睡覺，/你/進來/作/什麼」(稱自己)  
2) 「議論/人家的/好歹」(稱別人)

語素義和詞義分別指該語素或該詞在歷時語料庫中的語義。從古漢語發展到現代漢語，語素和詞都在不斷地發展變化，一個詞或語素可能會因此而有多個義項。有些詞的每一項詞義都存在逆序現象，有些只有部分詞義存在逆序，還有些詞和其對應的逆序詞完全異義。例如以上所述「人家」，和「家人」詞義不同，「人家」除了用作人稱代詞外，還有以下多項詞義。由於構詞語素「家<sub>3</sub>」與「家<sub>2</sub>」的語素義相同或相關，我們將「人家<sub>3</sub> — 家<sub>2</sub>人」作為一對同素逆序詞：



1) 《水滸傳》：「約/有/五七百/人家」(住戶)  
2) 《金瓶梅》：「他/是/詩禮/人家」(家庭)  
3) 《紅樓夢》：「因此/說/已/有/了/人家」(女子未來的夫家)

有些詞的詞義較特殊，不在本文進行分析：

1) 數詞：因數詞組合靈活，且沒有其他詞義，例如「一十一—十一」，因此不在本文分析。但如果其中一個詞的詞形為數詞，詞義具有引申義，則作為同素逆序詞研究，例如「一萬—萬一」。

2) 專有名詞：在建構歷時語料庫時，我們對人名、稱謂、官爵名、地名、建築名、機構名增加了命名實體標註，但鑒於這類詞的特殊性，不宜與普通詞彙混合在一起研究，因此不在本文進行分析。如有一詞多義的情況，例如「巡按」，既可用作官職名，也可指巡行按察，則只取其為普通名詞的情況進行研究。

3) 臨時製造的偶發詞：為了達到某種修辭效果或渲染具體的場景，文學作品中有時會臨時造詞，例如「不/偷/漢子/，/只/偷/子漢」，由於「子漢」是為了戲謔而臨時編造的詞，因此不納入本文的研究範圍之內。

#### 4 歷時語料庫中雙音同素逆序詞的分布

對語料庫詞條和同素逆序詞的統計僅包括第3節界定的範圍，一詞多義的情況則根據其詞義判斷，只取其在本文研究範圍所界定類別的數據。例如「千萬」，可為數詞，也可指「務必」，由於數詞不在本文的研究範圍內，因此對其用作數詞的情況不作統計。

<sup>1</sup> 據《現代漢語詞典》[13]、《普通話水平測試綱要》(馬顯彬，廣州：暨南大學出版社，2001)、《新華字典》(北京：商務印書館，1998)，「應酬」的「酬」讀輕聲，是普通話中的變調，台灣讀為第二聲(台灣教育部《重編國語辭典修訂本》)。

表 1 展示了各文本中雙音同素逆序詞的對數，不考慮詞頻。「雙音同素逆序詞對/詞條的比率」即用雙音同素逆序詞的對數除以語料庫中分詞之後的詞條數（不計重複），此比率顯示了雙音同素逆序詞的豐富度。總體看來，明清小說中的同素逆序詞較先秦典籍豐富得多，其原因是多方面的：1) 詞彙發展的因素：漢語詞彙的發展趨勢是從單音節向雙音節演進，單音節詞在先秦典籍中占更加主導的地位。2) 語體風格的因素：先秦典籍和明清小說的體裁差異較大，先秦典籍為史書，語言文雅，且其中的人物以貴族等官方人物為主，使用官話，語言比明清小說單純；而明清小說中的人物以市井小民為主，含較多方言俚語，語言多樣化成為其顯著特徵。3) 語言接觸的因素：兩部先秦典籍的內容寫及的範圍較小，當時的交通，人的流動範圍、頻率都更小，因此語言相互影響的機會更低。四部明清小說中，《水滸傳》的同素逆序詞最豐富，其中一個重要原因是其來源於民間文學，在成書之前已廣為流傳，因此保留了說唱藝術的語言特色，並夾雜大量方言口語，文風明快生動，因此用詞極為豐富。

表中所說的「同類體裁合併」是指分別將兩部先秦典籍、四部明清小說合併，考察同類體裁文本中的雙音同素逆序詞，將四部明清小說合併之後，雙音同素逆序詞多達 819 對，體現了這類詞對於研究明清小說語言特色的重要性。「同類體裁交集」是指分別將兩部先秦典籍、四部明清小說合併，考察同類體裁文本中共有的雙音同素逆序詞，兩部先秦典籍中有 6 對，四部明清小說中有 17 對，這些詞為該體裁中常見的同素逆序詞，例如先秦典籍中的「兵甲—甲兵」、「外內—內外」，明清小說中的「答應—應答」、「姓名—名姓」。「全部合併」是指將六部著作合併，總體考察其中的雙音同素逆序詞，這一數據有助於比較這些詞在不同時代、不同體裁文本中的使用傾向，例如先秦典籍中使用「氏姓」，而明清小說中則使用「姓氏」，反映了其語言的時代特徵。「兩種體裁交集」考察的是先秦典籍和明清小說中共有的雙音同素逆序詞，共 9 對，這一數據有助於考察不同時代文本中這類詞的沿用情況和詞義的演變，例如「甲兵」，《左傳》中的詞義「戰爭」，未見於明清小說中，辭源（修訂本）[12]和《現代漢語詞典》[13]中也未列出此義項，可見其詞義有所縮小。

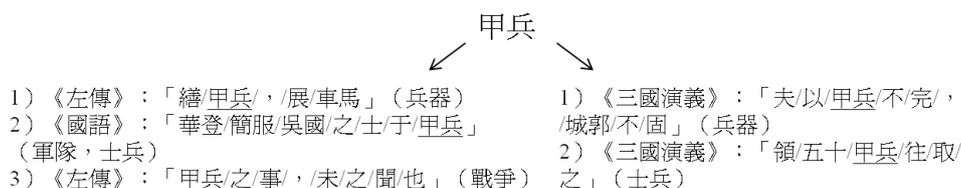


表 1：雙音同素逆序詞的分布（類型）

項目		先秦典籍		明清小說			
		左傳	國語	三國演義	水滸傳	金瓶梅	紅樓夢
分詞之後的詞條（條）		6,396	4,392	12,933	18,984	21,164	20,290
雙音同素逆序詞（對）		37	19	150	288	208	228
雙音同素逆序詞對/詞條的比率（%）		0.58%	0.43%	1.16%	1.52%	0.98%	1.12%
雙音同素逆序詞（對）	同類體裁合併	71		819			
	同類體裁交集	6		17			
	全部合併			956			
	兩種體裁交集			9			

表 2 展示了各文本中雙音同素逆序詞的使用頻率分布，表中關於不同文本合併、交集的計算方法和表 1 相同。從詞頻的角度考察雙音同素逆序詞的使用，有助於觀察其在不同時代、不同體裁文本中的使用傾向及發展、穩定趨勢。相比於表 1 對類型的統計數據，在同素逆序詞的使用頻率上，先秦典籍和明清小說的差異更大，兩部先秦典籍的「雙音同素逆序詞/詞條的比率」均不超過 0.20%，而《水滸傳》高達 1.69%，由此可見兩種體裁的語言風格區別明顯。將同類體裁合併之後，先秦典籍中使用頻率高低差異比最大的三對詞為「子弟/20—弟子/1」、「民人/21—人民/2」、「甥舅/10—舅甥/1」，其中第 1 對詞的詞義不同，第 2、3 對同義詞的詞頻差異顯示了先秦典籍的用詞傾向。明清小說中使用頻率高低差異比最大的三對詞為「吩咐

/765—吩咐/1」、「背後/468—後背/1」、「方纔/227—纔方/1」，其中第 2 對詞的詞義不同，第 1、3 對同義詞的詞頻差異巨大，該詞的取詞傾向，顯示了其正在接近現代漢語的趨勢。

表 2：雙音同素逆序詞的分布（詞頻）

項目	文本		明清小說			
	先秦典籍	國語	三國演義	水滸傳	金瓶梅	紅樓夢
分詞之後的詞條（次）	130,678	54,025	318,780	486,204	411,174	494,194
雙音同素逆序詞（次）	200	109	2,570	8,212	4,337	5,432
雙音同素逆序詞/詞條的比率（%）	0.15%	0.20%	0.81%	1.69%	1.05%	1.10%
雙音同素逆序詞（次）	同類體裁合併	459		28,461		
	同類體裁交集	128		6,094		
	全部合併			31,414		
	兩種體裁交集			2,632		

## 5 兩種體裁中雙音同素逆序詞的對比分析

通過對歷時語料庫的窮盡統計，可以了解各文本、各類體裁中同素逆序詞的分布概貌。在本節中，我們將兩部先秦典籍和四部明清小說分別合併，觀察這兩種體裁的文本中雙音同素逆序詞的使用情況，並基於數據和實例，從詞義的角度進行對比分析，其結果可用於考察這些詞的沿襲、淘汰和演變。根據兩種體裁中同素逆序詞的使用情況，可分為以下幾類：

1) 先秦典籍和明清小說中共有的詞對：由於這兩類體裁文本的語體風格差異較大，因此共有的同素逆序詞僅 9 對，均為上古漢語和近代漢語中都會使用的常見詞，從以下幾對詞義相同的同素逆序詞在詞頻上的差異可見這些詞彙的發展和穩固情況：

短長/（先秦典籍 1，明清小說 1）—長短/（先秦典籍 1，明清小說 26）

職官/（先秦典籍 1，明清小說 1）—官職/（先秦典籍 5，明清小說 11）

根據上下文語境的不同，一詞多義的情況較普遍，有時逆序現象只存在於某一種語義之下，而並不存在於該詞彙的其他語義。例如「職官」在先秦典籍中有兩項詞義：a) 官職（《左傳》「先父之/職官/也」）；b) 泛指眾官（《左傳》「職官/五正」），在考察詞義相同的同素逆序詞時，只統計「職官」的詞義為「官職」的數據。

2) 僅見於先秦典籍：共 29 對，需要說明的是，其中有些詞雖未見於四部明清小說，但在近現代漢語中仍在使。以下幾對同義詞的詞頻數據顯示了這些詞彙在同時代史籍中的使用傾向：

胄裔/1—裔胄/1      惠慈/1—慈惠/3      侈驕/1—驕侈/2

3) 僅見於明清小說：共 735 對，以下幾對同義詞的詞頻差異說明了一個現象，即明清小說的語言介於古、中漢語與現代漢語之間，在文言文與語體文之間扮演了一個過渡的角色：

據占/1—占據/18      迫逼/1—逼迫/14      鬧熱/4—熱鬧/124

4) 先秦典籍中僅用 A，明清小說中僅用與其相對應的同素逆序詞 B：共 76 對，從詞義對比的角，可大體分為以下幾類：

a) A、B 詞義相同或相近：對這類詞進行比較分析可以管窺其演變趨勢，明清小說中的很多詞彙已接近現代漢語，例如：

禍災（先秦典籍）—災禍（明清小說）      寶財（先秦典籍）—財寶（明清小說）

b) A、B 詞義不同：

弄戲—戲弄

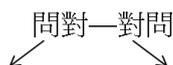
《國語》：「夷吾之/少/也，/不好/弄戲」（遊戲）      1) 《紅樓夢》：「誰敢戲弄/你」（捉弄）  
2) 《金瓶梅》：「向/前/戲弄/那孩兒」（逗引）

c) A 為本義，B 為引申義：

雪霜—霜雪

《左傳》：「則/雪霜/風雨/之/不時」（本義）      《三國演義》：「志/懷/霜雪」（比喻高潔的情操）

d) A、B 詞義相近，但情感色彩不同：



《左傳》：「齊侯問對於晏桓子」（一問一答，中性） 《水滸傳》：「去/拿/宋江/來/對問」（審問，含貶義）

## 6 結語

本工作從自然語言處理的角度出發，基於包含兩部先秦典籍和四部明清小說的歷時語料庫，對其中的雙音節同素逆序詞進行比較研究，該語料庫已完成分詞和命名實體標註，但尚未加上詞性標註。根據對歷時語料庫中雙音同素逆序詞的窮盡統計，其結果顯示，無論是從同素逆序詞的豐富度還是使用頻度來看，明清小說都遠遠高於先秦典籍，其原因是多方面的，包括詞彙從單音節向雙音節發展的趨勢，兩類體裁的語體風格差異，及語言接觸的廣度。本文還進一步對這兩種不同體裁文本中雙音同素逆序詞的使用情況進行比較分析，其結果有助於考察這些詞在不同體裁文本中的使用傾向，及在不同時代的沿用、淘汰和變遷。因一對同素逆序詞的詞性可能有所不同，有些詞還具有多種詞性，例如「擔負—負擔」，前者只能作動詞，而後者可作動詞和名詞。在後續的工作中，將為語料庫增加詞性標註，從詞性和語法功能的角度進一步進行比較研究，還將引入異體字對應知識，以增加數據分析的完整性。

## 參 考 文 獻

- [1] Xiong D, Lu Q, Lo FJ, Shi DX, Chiu TS. Specification for Segmentation and Named Entity Annotation of Chinese Classics in the Ming and Qing Dynasties. In: Chinese Lexical Semantics, Lecture Notes in Computer Science, Volume 7717. Berlin, Heidelberg: Springer, 2013. 280~293.
- [2] 姜黎黎。古代漢語同素異序詞研究綜述。江南大學學報（人文社會科學版），2009，8（3）：108~112。
- [3] 楊芬。同素異序詞研究現狀綜述。科技創新導報，2010年第6期：10~11。
- [4] 張巍。中古漢語同素逆序詞演變研究。復旦大學博士學位論文，2005。
- [5] 張德鑫。談顛倒詞。漢語學習，1995年第06期：18~22。
- [6] 張能甫。東漢語料及同素異序的時代問題。古漢語研究，2000年第03期：74~77。
- [7] 譚汝為。同素逆序詞四論。詞彙學新研究—首屆全國現代漢語詞彙學術研討會選集。北京：語文出版社，1995：146~161。
- [8] 艾鳳美。關於同素異序詞的界定。語文學刊，2008年第24期：118~126。
- [9] 趙艷。基於《現代漢語逆序詞典》的同素逆序異義詞研究。華中師範大學碩士論文，2013。
- [10] 漢語大字典編輯委員會。漢語大字典。武漢：湖北辭書出版社，成都：四川辭書出版社，1986~1990。
- [11] 羅竹風。漢語大詞典。上海：上海辭書出版社，2007。
- [12] 廣東、廣西、湖南、河南辭源修訂組。辭源（修訂本）。北京：商務印書館，1998。
- [13] 中國社會科學院語言研究所詞典編輯室。現代漢語詞典（第6版）。北京：商務印書館，2012。
- [14] 楊伯峻、徐提。春秋左傳詞典。北京：中華書局，1985。
- [15] 毛遠明。左傳詞彙研究。重慶：西南師範大學出版社，1999。
- [16] 陳長書。《國語》詞彙研究。山東大學博士學位論文，2005。
- [17] 倪懷慶。《國語》詞彙研究。廣州大學碩士學位論文，2004。
- [18] 沈伯俊、譚良嘯。三國演義大辭典。北京：中華書局，2007。
- [19] 李法白、劉鏡芙。水滸語詞典。上海：上海辭書出版社，1989。
- [20] 胡竹安。水滸詞典。上海：漢語大詞典出版社，1989。
- [21] 白維國。金瓶梅詞典。北京：中華書局，1991。
- [22] 黃霖。金瓶梅大辭典。成都：巴蜀書社，1991。
- [23] 周汝昌。紅樓夢辭典。廣州：廣東人民出版社，1987。
- [24] 馮其庸、李希凡。紅樓夢大辭典。北京：文化藝術出版社，1990。
- [25] 俞士汶、段慧明、朱學鋒、孫斌、常寶寶。北大語料庫加工規範：切分·詞性標註·注音。Journal of Chinese Language and Computing, 2003, 13 (2): 121~158。
- [26] 中文資訊處理分詞規範。臺灣經濟部中央標準局印行。
- [27] 劉源，等。資訊處理用現代漢語分詞規範及自動分詞方法（中國國家標準 GB13715 信息處理用現代漢語分詞規範）。北京：清華大學出版社，1994。