

基於語料庫的學習型漢語詞典例證搭配研究

陽捷*, 孫建榮

(澳門科技大學國際學院, 澳門)

摘要: 本研究以語料庫中學習者的近義詞(“產生”和“發生”)使用情況以及學習型詞典中近義詞(“產生”和“發生”)的例證搭配為文本, 分析其在兩種環境中搭配行為之特徵, 並進行對比, 以管窺學習型漢語詞典在例證收錄和呈現方面的特點。分析結果發現學習型漢語詞典在處理例證信息時, 有待在搭配詞語典程度、搭配信息充分性以及詞義辨析方面進一步改善。在此分析結果上, 本研究提出了可考慮的解決方式。

關鍵詞: 近義詞; 詞典例證; 語料庫; 典型搭配

A Corpus-based Study on Collocation of Examples in Chinese Dictionaries for Foreign Learners

Jie YANG*, Jianrong SUN

(University International College, Macau University of Science and Technology, Macau, China)

Abstract: This paper analyzes the features manifested in the usage of the Chinese synonyms of “Chansheng” (“to emerge”) and “Fasheng” (“to occur”) in language use data corpus and the explanatory information as illustrated in usage examples provided in dictionaries for language learners. A comparison is made of the two types of analyses. The analyses reveal some discrepancies between usage examples and use authenticity in that usage examples provided in dictionaries lack representativeness and sufficiency in connotation meanings. Error analyses of learners’ use of the pair of the synonyms in Chinese suggest that an absence of considering factors concerning the learners might lead to inadequate use of the Chinese synonyms.

Keywords: Synonyms; Dictionary examples; Corpus; Collocation

0 引言

近義詞, 是指詞彙意義相近的詞語, 如果兩個詞語“近義”, 意味著它們的意義存在部分一致的反映內容, 但同時彼此的意義必有不同的外延^[1]。近義詞的區分, 是漢語二語教學的一個難點, 在口語和書面語產出時, 學習者往往會因為沒有全面瞭解其區別而出現偏誤。以“產生”和

“發生”為例, 在《漢語水準詞匯與漢字等級大綱》中, “發生”屬於甲級詞, “產生”屬於乙級詞; 在 2010 年國家語言文字工作委員會發布的《漢語國際教育用音節漢字詞匯等級劃分》中, “產生”和“發生”都屬於一級(初級)詞匯, 一級(初級)詞匯被稱為普及化等級, 是現代漢語常用詞。這兩個漢字詞匯等級大綱都是基於詞頻劃分詞匯等級, 因此可以看出, 在實際應用中, “產生”

收稿日期: 2021-03-03; 修訂日期: 2021-03-24。

* 通訊作者: 陽捷, 女, 在讀博士, 主要研究方向: 應用語言學。

E-mail: 1909853mut30001@student.must.edu.mo, Tel: +86-18236517603

和“發生”使用頻率的確很高。再來看漢語教學，在國際漢語教學中，這一組詞，即“產生”和“發生”，作為近義詞，常被漢語學習者混淆和誤用。以 HSK 動態作文語料庫（下文簡稱 HSK 語料庫）的檢索為例，在該語料庫統計分析中，可以看到，“產生”和“發生”這組近義詞，僅在錯詞這一項有關詞語錯誤的匯總中，錯誤頻次就分別達到 139 和 189 次，居於錯誤頻次統計的前列。這些誤用，主要是詞匯的搭配誤用。詞匯搭配是學習者詞匯學習一個不可忽視的範疇，詞匯學習的方式和渠道非常多，本研究則主要從學習型詞典的角度出發，來探討工具書在漢語學習者詞匯搭配學習中的重要示範和指導作用。

作為詞彙學習和辨析的重要參考，學習型詞典，即為非本族語者學習目標語而編撰的詞典^[2]，在學習者的語言學習過程中，扮演著不可或缺的重要角色。在學習搭配，產出目標語的過程中，詞典例證（例證是詞典中釋義的補充和延伸，為詞目提供具體的實例）所展示的詞語搭配等用法資訊，對學習者理解和產出目標語，起到關鍵的參考和示範作用。這一點也可以從學習者的語言產出中，窺見一斑。

以近義詞“產生”和“發生”為例，在 HSK 語料庫中，檢索“產生”和“發生”的搭配，會發現“問題”“矛盾”等出現在學習型詞典例證搭配中的詞語，往往是學習者產出的高頻搭配詞語。另外值得注意的是，也正因為學習型詞典的例證相關問題，有可能會造成學習者在語言產出時出現偏誤。例如，在《商務館學漢語》詞典中，近義詞“產生”和“發生”都可以搭配“矛盾”，表示某個事情或事物出現了。詞典所列例證為“他們之間好像發生（產生）了矛盾”。而以“產生”或“發生”為檢索詞，考察 HSK 語料庫，會發現學習者產出，諸如“這一鴻溝阻礙了各輩之間通暢的溝通，使得他們之間發生了很多矛盾”等錯誤搭配的句子。另外如《現代漢語學習詞典》中，“發生”一詞，列出了——“發生問題”這一例證，則相對應可以在 HSK 語料庫中，檢索到“我們常

常看到因代溝發生的許多家庭問題”、“代溝問題就是因兩個不同年代的人互相不溝通而發生的一系列的問題”等錯誤搭配的句子，並且這些搭配錯誤的句子所占比例很高（下文有詳細統計）。那麼學習者在近義詞學習和辨析過程中，出現的這些高頻誤用，是否與詞典的例證有聯繫？或者說，學習型詞典的例證是否考慮了學習者近義詞學習的相關特點，起到了其作為規範和參考的作用？相關的問題，需要基於數據的分析和探討，而以往關於近義詞的研究表明：採用語料庫方法，對近義詞進行研究，能夠幫助學習者有效區分近義詞^[3-4]。因此本文採取基於語料庫的研究方法，以“產生”和“發生”這一組近義詞為考察對象，探究其在學習型詞典例證搭配方面的問題。

1 相關研究綜述

採用語料庫進行近義詞辨析方面的研究，主要集中在英語方面。如 Biber 等對近義詞的用法、語境、語域進行了分析^[3]；Kennedy 借助語料庫，對近義詞“between”和“through”進行了區分^[4]；Thomas 則辨析了詞語“keep”和“get”等等^[5]。在國內的英語研究方面，基於語料庫的近義詞辨析和教學方面的相關研究，數量也不少，如王春豔，從句法、語義韻和非詞語化用法等方面區分了“true”和“real”這組近義詞的不同^[6]；陸軍通過分析，指出學習者英語近義詞的使用特徵和發展模式與他們對近義詞搭配行為和語義韻知識掌握的欠缺相關^[7]。從以往的對於英語近義詞的研究可以看出，近義詞的區分和辨析，離不開對其的搭配的探索。研究者對目標詞語的搭配行為，類聯接和語義傾向的關注，在本質上都離不開詞語在具體或抽象層面的搭配行為。詞語搭配指的是詞與詞的結伴使用，是詞語間的典型共現行為^[8]。語料庫研究先驅，John Sinclair 指出，短語是意義承載的最重要單位^[9]。這個“短語”實際上聚焦的就是詞語的形式、意義和功能的共選行為特徵^[10]。

從搭配層面切入，對於英語近義詞的大量實證

研究，也給對外漢語近義詞的辨析和教學，帶來了借鑒和參考。陸方喆通過研究，指出詞語搭配是對外漢語近義詞辨析的一個重要層面^[11]；曾利，陳從梅對“提升”在語法層面的搭配組合關係，進行了考察，嘗試構建漢英學習型詞典中動詞詞目的類聯接信息表徵方式^[12]；曹蓉則採用語料庫驅動的語言研究方法，對現代漢語“一起”的義項進行了研究，研究指出以大規模真實使用作為第一證據的語料庫語言研究方法，在詞典釋義方面，能為語言學習者提供一種意義辨識的優先級參照^[13]。相關的研究成果，為學習者學習漢語近義詞提供了思路。學習者對近義詞的掌握，離不開對其典型搭配的習得。而對典型搭配的習得，也離不開對詞典，尤其是學習型詞典例證的參考。因為例證“集語義、語用、語法於一身，能讓讀者從句子的整體上掌握相關的語言現象”^[14]。而且學習詞典中的例證如同教材中的範句，應內容合理，不能含混不清^[15]。研究者們對有關例證的編寫原則和理論指導方面的關注，也推進了詞典例證在宏觀和微觀層面的研究。劉曉梅、梁青通過對《商務館學漢語詞典》例證的文化傳播功能的評析，指出漢語學習詞典的編纂應該注意發揮例證的文化傳播功能^[16]；張宏則通過對漢外詞典中常用動詞的例證研究，指出例證配置存在的問題，認為給不同詞類制定相應的配例方案，是解決例證問題的關鍵^[17]；劉慧以“可想而知”為例，對比分析了四部漢語學習詞典，提出利用語義韻研究改進詞典釋義及用例，並指出語料庫語言學的研究方法是對外漢語學習詞典的努力方向之一^[18]。雖然基於實際考察的例證研究，越來越受到研究者的關注，但數量還相當有限，更多的只是研究者在對詞典文本進行整體描寫，評介和評估時，進行的局部分析。而對於具體的類別的詞語的實證研究，比如近義詞，還有待進一步拓展。近義詞的學習是漢語學習的一個難點，學習者容易誤用搭配，產生偏誤。學習者進行近義詞的學習時，往往離不開對詞典例證的參考。例證是他們理解漢語，產出漢語的重要參照對象。因此學習型詞典的近義詞用例要充分考慮讀者對象的特殊需要。也就

是說，在推進近義詞教與學的過程中，學習型詞典除了體現漢語本身的相關典型特徵，還應結合學習者特殊學習需要，在例證上體現出“外向”特徵。本文在前人研究的基礎上，利用現代漢語平衡語料庫和 HSK 動態作文語料庫，選取“產生”和“發生”為研究對象，一方面通過擴大語料廣度增加對這對常見近義詞的辨析，另一方面通過對比分析詞典例證與本族語真實語料、學習者真實語料，在搭配層面的差異，探究有針對性且有效的近義詞詞語搭配，為相關例證的選取和編撰提供參考。

2 研究工具和步驟

2.1 研究工具和對象

本文主要通過考察真實語言實例中近義詞使用的搭配行為特徵，來探討漢語學習型詞典的相關例證現狀。詞典選取的是《商務館學漢語》和《現代漢語學習詞典》。選取這兩本詞典為研究對象，主要是因為其比較具有代表性，是漢語教師與學生所常用的學漢語工具。另外研究主要借助國家語委現代漢語語料庫和 HSK 動態作文語料庫平臺所提供的資料。國家語委現代漢語語料庫，是一個大型、通用的現代漢語平衡語料庫，語料選材類別廣泛，時間跨度大，旨在服務於學術研究、語言教育、資訊處理，標準規範制定。其語料庫字元數為 19455328 個。HSK 動態作文語料庫，是母語非漢語的外國人參加高等漢語水準考試（HSK 高等）作文考試的答卷語料庫，語料庫約為 424 萬字。為了行文方便，在下文中，國家語委現代漢語語料庫簡稱為現漢語料庫，HSK 動態作文語料庫簡稱為 HSK 語料庫。

2.2 研究步驟

1) 《商務館學漢語》和《現代漢語學習》詞典中“產生”和“發生”的例證及典型搭配分析。

2) 學習者語料 (HSK 語料庫) 中“產生”和“發生”的搭配使用考察。

3) 本族語語料庫 (現漢語料庫) 中“產生”和“發生”的搭配考察。

4) 對比分析“產生”和“發生”在詞典例證和語料庫中的搭配信息, 討論並總結。

既包括具體層面的詞語搭配, 還包括抽象層面的詞語搭配——類聯接。類聯接是詞語搭配發生於其中的語法結構和框架。“一個類聯接代表了一個搭配類。比如, 人們常說的 N+V 就是一個類聯接, 它代表一類搭配, 在研究中稱為 N+V 搭配”^[20]。

3 數據和討論

詞語之間的組合關係, 線性共現, 體現出詞語搭配^[19]。而詞語的組合必然涉及句法, 即詞語在語法範疇間的結合。因此本文所探討的搭配,

3.1 “產生”和“發生”的例證及典型搭配分析

《商務館學漢語》和《現代漢語學習詞典》中“產生”和“發生”的例證列表如下:

表 1 詞典中“產生”和“發生”的例證

詞典	“產生”例證 (按詞典例證給出順序)	“發生”例證 (按詞典例證給出順序)
《商務館學漢語詞典》	產生社會矛盾 / 就業問題是怎樣產生的呢? / 會上產生了兩種意見 / 常務委員會由委員會全體委員選舉產生 / 對漢字產生了興趣 / 他們之間好像產生了矛盾 / 產生了新的領導集體 / 這次活動產生了很好的效果	發生事故 / 發生變化 / 事情發生在上午九點 / 昨天這裡發生了一起交通事故 / 故事發生在很久以前 / 村子裡從來沒發生過這樣的事 / 事情發生得太突然, 我們都沒想到 / 對漢字發生了興趣 / 他們之間好像發生了矛盾 / 發生了地震 / 昨天這兒發生了一次交通事故
《現代漢語學習詞典》	語言中產生了一些新的詞彙 / 新時代產生新思想 / 他又產生了實現理想的希望 / 這樣的好人好事, 只有在新時代才會產生 / 他對唱京戲產生了濃厚的興趣 / 大家勸他的話沒有產生一點作用 / 產生新的領導班子	發生問題 / 發生事故 / 這種情況以前也發生過 / 發生關係 / 發生衝突 / 發生摩擦 / 發生爭吵 / 他對唱京戲發生了濃厚的興趣 / 大家勸他的話沒有發生一點兒作用 / 發生意外事故

根據對“產生”和“發生”的語法信息的研究，經過分析，“產生”和“發生”的搭配列表如下：

表 2 詞典中“產生”和“發生”的搭配列表

	《商務館》產生	《商務館》發生	《現漢學習》產生	《現漢學習》發生
詞匯搭配	名詞 / 名詞性方位短語 +~	【主調】	【主調】	【主調】
	問題~	事情~	語言中~	情況~
	他們之間~	這裡~	新時代~	話~
	會上~	故事~	他~	
	活動~	村子裡~	話~	
		他們之間~		
		這兒~		
	~+ 名詞 / 動詞	【動賓】	【動賓】	【動賓】
	~矛盾	~事故	~詞彙	~問題
	~意見	~變化	~思想	~事故
	~興趣	~事	~希望	~關係
	~矛盾	~興趣	~興趣	~衝突
	~集體	~矛盾	~作用	~摩擦
	~效果	~地震	~班子	~爭吵
				~興趣
				~作用
	~+ 介詞短語	【動補】		
		~在上午九點		
		~在很久以前		
	~+ 副詞	【動補】		
		~得突然		
	介詞短語 / 副詞 +~	【狀中】	【狀中】	【狀中】
	由...~	對...~	在...~	對...~
	對...~	好像~	對...~	
	好像~			

從表 2 中可以看出，《商務館學漢語》和《現代漢語學習詞典》中“產生”和“發生”的搭配主要是三類。一類是和名詞或名詞性方位短語形成主調結構；另一類是和名詞或動詞形成動賓結構；第三類是和介詞短語或副詞形成狀中結構。在詞典呈現的整個目標詞的搭配中，可以從詞語搭配出現的頻數看出，動賓結構所占的比例最大，其次是主調結構。“產生”和“發生”是傳統意義上的近義詞，因此為了辨析這組詞，兩部詞典在搭配方面，除了體現出“產生”和“發生”在搭配方面的不同之外，也列出了可以替換的搭配

詞語，如《商務館學漢語》中，“他們之間...”“興趣”“矛盾”“對...~”“好像”，以及《現代漢語學習詞典》中，“話”“興趣”“作用”“對...~”。另外《商務館學漢語》中，還呈現了“發生”一詞和介詞短語以及副詞構成的動補結構。詞典例證的目的主要是使讀者進一步理解詞義，瞭解詞的用法^[21]。在二語學習中，研究者通過對漢語學習者的調查，發現大部分學習者在遇到近義詞時，會在詞典中的典型搭配和例句中尋找辨析信息^[22]。那麼詞典例證對學習者近義詞詞語產出的作用如何，可以通過對真實語料的考察窺見一斑。

3.2 學習者語料（HSK語料庫）中“產生”和“發生”的搭配使用考察

考察學習者的語言產出，可以瞭解其對相關目標詞語的辨析和掌握。而對其中出現的相關誤用的分析，能進一步瞭解學習者在近義詞辨析方面的重點和難點，同時對詞典例證的收錄和編排也起到提示作用，使之更好地體現出“外向”特徵。因為目前的學習詞典“用例內容和語言沒有完全跳出面向國人的詞典的路子，沒有充分考慮

讀者對象的特殊需要”^[23]。

在 HSK 語料庫中，分別檢索節點詞“產生”和“發生”，得到含有“產生”的索引行 1450 條，“發生”的索引行 1644 條。經過統計發現，與“產生”相關的在搭配層面的錯誤頻次為 461，錯誤比例約為 32%；與“發生”相關的在搭配層面的錯誤頻次為 393，錯誤比例約為 24%。對學習者節點詞“產生”和“發生”的主要搭配誤用分析（因篇幅原因，按降冪排列出典型的搭配誤用），列表如下：

表 3 “產生”的典型搭配偏誤

排序	結構	“產生”的主要搭配偏誤	出現的頻數	所占比例	搭配偏誤總頻數
1	動賓	“生產”與“產生”混淆 賓語前的動詞應為“生產”，誤用成“產生”	88	19%	461
2	動賓	“產生”與“發生”混淆 賓語前的動詞應為“產生”，誤用成“發生”	69	15%	
3	動賓	賓語前缺少動詞“產生”	41	9%	
4	狀中/ 動補	“產生”前面缺少情態動詞“會” 或後面缺少助詞“了”等	38	8%	
5	動賓	“產生”與其它近義詞混淆賓語前的動詞應為 “產生”，誤用成其它動詞 (“出來”“導致”“感到”等)	30	7%	
6	動賓	其它動詞與“產生”混淆 賓語前的動詞應為其它動詞 (如“帶來”“形成”“造成”等)	15	3%	

表 4 “發生”的典型搭配偏誤

排序	結構	“產生”的主要搭配偏誤	出現的頻數	所占比例	搭配偏誤總頻數
1	動賓	“產生”與“發生”混淆 賓語前的動詞應為“產生”，誤用成“發生”	97	25%	393
2	動賓	其它動詞與“發生”混淆 賓語前的動詞應為其它動詞[（如“得（病）” “出現”“發出”“引起”等]	61	16%	
3	狀中	“發生”前面缺少情態動詞“會”	23	6%	
4	動賓	賓語前缺少動詞“發生”	19	5%	
5	動態助詞 “了”	“了”缺少或多餘 “發生”後多詞“了”或缺詞“了”	24	6%	

觀察表 3 和表 4 可以發現，出現誤用最高的還是動賓結構的搭配，這也跟“產生”“發生”的詞性相關。“產生”的典型搭配偏誤中，排在首位的是“生產”與“產生”的混淆，出現的錯誤搭配如：

a. 如果農民們都要產生綠色食品，農作物的產量會大大降低。

b. 但是，由於世界上的人越來越多，糧食的需求更高，農民們就得按最可能快的辦法產生很多糧食。

c. 但如果人們為了不挨餓產生這樣的農作物的話，他們的身體會越來越惡化。

“產生”和“生產”在形態上，只是順序有差別，然而在詞義和搭配上卻大不相同。“生產”除了用作生孩子外，更常見的用法是搭配具體的生產資料或生活資料的名詞，如“食品”、“糧食”、“物品”等；“產生”更常見的是搭配抽象義的名詞，如“矛盾”“意見”“興趣”等。這組詞不是傳統意義上的近義詞，屬於“非典型”近義詞，對於本族語者來說，很容易將其區別開來，但是漢語學習者對其的辨析卻存在困難，因此詞典在例證辨析時，應從學習者的特殊需求出發，擴大近義詞辨析的範圍。

另外，“產生”與“發生”的混淆，在“產生”的搭配誤用中排在第二位，在“發生”的搭配誤用中則排在首位。這同時也說明，這組詞的搭配區分，對於學習者來說是一個難點。通過檢索分析，發現這一類別的搭配中排在前三位的錯誤搭配詞語有“問題”（出現頻數 28）、“代溝”（出現頻數 19）、“矛盾”（出現頻數 15）。“問題”“代溝”“矛盾”前的動詞應為“產生”，但誤用成“發生”，如：

a. 青年人的學習速度比老年人還快的情況下，必然發生溝通的問題。

b. 現代也是如此，在兩代人之間發生的“代溝”，讓家庭成員之間產生許多問題。

c. 這一鴻溝阻礙了各輩之間通常的溝通，使得他們之間發生了很多矛盾。

對比《商務館學漢語》和《現代漢語學習詞典》可以發現，出現在詞典例證中的節點詞的搭配，是學習者較常使用的搭配。但同時也是學習者容易發生錯誤的搭配。從表 2 中，可以看出，《商務館學漢語》中出現了主謂結構的“問題產生”（就業問題是怎樣產生的呢？），而《現代漢語學習詞典》中出現了動賓結構的“發生問題”（發生問題），也就是說“問題”既可以和“產生”，也可以和“發生”進行搭配。但從括弧中給出的例證可以看出，詞典並未給出能夠幫助學習者辨析的搭配例證。同樣，以“矛盾”為例，《商務館學漢語》中，提到表示某個事物或事情出現了時，有時“產生”和“發生”可以換用，並給出例證——“他們之間好像產生/發生了矛盾”。但從上文的例句 C，可知“發生了很多矛盾”也是某個事物或事情出現，但更恰當的搭配應該選用“產生”。此類的例子可以從 HSK 語料庫中檢索到，如：

a. 過去，爸爸和媽媽不理解孩子，孩子這一輩覺得父母們的想法和觀念太古老了。所以，父母和兒女之間常常發生矛盾。

b. 但是在現代家庭裡，所有家人都很忙，幾乎沒有說話的時間，當然互相瞭解得不太深，因此往往發生家人之間的矛盾。

c. 過去，因代溝而發生的矛盾不太多。因為孩子聽從父母的話是天經地義的事情。

以上的“矛盾”，以及詞典中提到的“興趣”、“作用”等“產生”和“發生”可以換用的搭配，如果以詞典中給出的例證為參考，那麼學習者往往會在搭配方面出現偏誤。也就是說詞典在給出例證時，尤其是可以換用的近義詞的搭配時，應該對具體的搭配加以說明，並給出典型的語境。例如，如果是有原因，由某個行為或事件導致，更恰當的搭配應該選用“產生”而不是“發生”。這一點也可以聯繫“產生”的“產”的本義，“產”的本義為出生、生育，即從已有中產生新有，有前因後果。

另外，在學習者的典型搭配偏誤中，還出現了賓語前缺少動詞“產生”，以及“產生”“發生”

前缺少情態動詞“會”“能”等偏誤。對比分析詞典可以發現，節點詞“產生”和“發生”的例證中，並沒有出現這類典型的狀中結構。其它的典型偏誤，也進一步說明，學習者對節點詞“產生”、“發生”的典型搭配並未區分開來。那麼“產生”和“發生”的典型搭配有哪些呢？因為就搭配而言，在語料庫研究先驅 John Sinclair 看來，任何搭配都是可能的，只不過一些比另一些更為恰當^[24]。恰當的搭配，是指詞語搭配要具有典型性，要考慮搭配的概率屬性。對恰當的詞語搭配的提取是詞彙搭配研究的一個重要方面。

3.3 本族語語料庫（現漢語料庫）中“產生”和“發生”的搭配考察

詞彙搭配研究中，有多種典型詞語搭配抽取方法，本文採用 antconc 軟件提取詞語搭配，並通過 MI 值詞彙搭配統計測量 (Mutual Information Score)，MI 值又稱為互信息值，表示的是一個詞在語料庫中出現的頻數所能提供的關於另一個詞出現的概率資訊)，MI 值計算公式為：

$$I(a,b) = \log \frac{w \times F(a,b)}{F(a) \times F(b)}$$

其中 W 是語料庫的總詞容量，F(a) 為詞形 a 的觀察頻數，F(b) 為詞形 b 的觀察頻數，F(a, b) 為兩個詞形在語料庫中的共現頻數。MI 值越大，說明節點詞對其共現詞吸引力越強。一般來說，在語料庫詞彙搭配研究中，MI 值等於或大於 3，就被認為是顯著搭配詞。MI 值能較好地識別複合詞，固定短語、科技術語^[25]。但是，也有可能因為語言使用者的個人用語特點或某個語料庫自身獨有特點出現些許偏差。另外在進行詞語搭配提取前，還需要對跨距，也就是節點詞的左右語境，進行界定。英語詞語搭配研究，一般將跨距界定為 -4/+4 或 -5/+5。而在漢語方面，孫宏林通過搭配自動抽取的對比試驗，指出漢語名詞、動詞、形容詞抽取搭配的最佳跨距分別為 [-2, +1]，[-3, +4]，[-1, +2]^[26]。本研究的跨距界定為 -3—+4，在提取搭配時，搭

配詞選擇條件統一定為 MI 值 >3。節點詞“產生”和“發生”的原始有效頻數分別為 5563 和 6317。本文分別以“產生”和“發生”的原始有效頻數為依據，考察其詞語搭配行為特徵。

節點詞“產生”和“發生”的左搭配詞的跨距界定為 -3/-1，右搭配詞的跨距界定為 -1/-4，與節點詞共現次數界定為 5。提取搭配之後，得出的詞通過定性分析剔除具有獨特學科特色的詞，以及不合語法和偶然落入跨距內的修飾詞。最後將“產生”和“發生”的典型左搭配詞和右搭配詞按降冪排列如下（表 5、表 6）：

表 5 “產生”的典型左搭配

排序	頻數	頻數 (左)	頻數 (右)	MI 值	搭配詞
1	7	7	0	4.78620	由 / 經…
2	11	11	0	4.66067	…時會
3	5	5	0	4.52317	…時能
4	9	9	0	4.49670	憑空
5	7	7	0	4.42363	一經
6	37	37	0	4.32323	由此
7	5	5	0	4.30078	賴以
8	6	6	0	4.20124	在…心裡
9	7	7	0	4.13413	便會
10	6	6	0	4.04924	將會
11	7	7	0	4.00859	…時所
12	9	9	0	3.95613	相結合
13	12	12	0	3.91173	之所以
14	248	248	0	3.84558	會
15	14	10	4	3.83867	靈感
16	10	9	1	3.78620	摩擦
17	5	5	0	3.78620	很難
18	17	16	1	3.78620	使人
19	18	16	2	3.74667	文字
20	39	39	0	3.7929	容易

表 6 “產生”的典型右搭配

排序	頻數	頻數(左)	頻數(右)	MI 值	搭配詞
1	14	0	14	4.68667	效應
2	8	0	8	4.61628	疑問
3	12	0	12	4.56381	影響
4	6	0	6	4.56381	歧義
5	8	0	8	4.46427	動搖
6	7	2	5	4.42363	意義
7	25	0	25	4.38566	後果
8	59	0	59	4.38344	熱量
9	15	0	15	4.37116	抗藥性
10	14	1	13	4.34563	懷疑
11	8	0	8	4.32677	興趣
12	5	0	5	4.30078	出
13	14	0	14	4.20124	飛躍
14	6	0	6	4.20124	想法
15	6	1	5	4.20124	不了
16	17	1	16	4.11878	幻覺
17	17	1	16	4.06631	後代
18	26	0	26	3.99479	(的)根源
19	15	0	15	3.99265	(n+)感
20	240	2	238	3.94490	影響

表 7 “發生”的典型左搭配

排序	頻數	頻數(左)	頻數(右)	MI 值	搭配詞
1	10	10	0	4.78257	新近
2	6	6	0	4.78257	將要
3	7	6	1	4.58993	戰事
4	5	5	0	4.51954	極易
5	5	5	0	4.51954	現象
6	5	5	0	4.51954	會
7	8	8	0	4.32314	亦可
8	16	16	0	4.32314	事變
9	5	5	0	4.29714	疾病
10	7	7	0	4.13049	…時所
11	19	19	0	4.12361	面貌
12	5	5	0	4.10450	眼前
13	13	13	0	4.09069	偶然
14	9	9	0	4.04561	剛才
15	34	30	4	3.98739	突然
16	8	8	0	3.97522	究竟
17	51	43	8	3.94720	疾病
18	5	5	0	3.93457	即可
19	59	59	0	3.89703	容易
20	35	35	0	3.88949	經常

表 8 “發生”的典型右搭配

排序	頻數	頻數(左)	頻數(右)	MI 值	搭配詞
1	5	0	5	4.78257	變革
2	6	0	6	4.78257	變化
3	10	0	10	4.78257	短路
4	6	0	6	4.78257	性關係
6	6	0	6	4.78257	共鳴
7	6	0	6	4.78257	(了/過+數詞)次
8	12	0	12	4.66709	虧損
9	9	0	9	4.63057	法律效力
10	9	2	7	4.63057	變故
11	7	0	7	4.58993	口角
12	13	0	13	4.57612	故障
13	6	0	0	4.56018	反應
13	6	0	6	4.56018	衝突
14	35	0	35	4.51954	形變
15	5	1	4	4.51954	爭議
16	20	1	19	4.46064	變異
17	7	0	7	4.42000	動亂
18	10	0	10	4.40406	火災
19	6	0	6	4.3673	興趣
20	14	1	13	4.3200	糾紛

根據以上表中“產生”和“發生”的類聯接信息，經過分析，列出該組近義詞的搭配（表 9）：

表 9 中可以觀察到，“產生”和“發生”的左搭配詞主要有：名詞、副詞、介詞短語、助動詞等，右搭配則主要是名詞和助詞。在副詞搭配方面，“產生”和“發生”都可以搭配“容易”“很難”“極易”等程度副詞。但“產生”前還出現了“憑空”“一經”等典型搭配詞，這與“產生”強調現象或事件原來沒有，現在出現了，側重前因後果有關。“發生”前較典型的副詞有“偶然”“剛才”“突然”等，這也跟“發生”側重事情發生無預期，出乎意料有關。左搭配名詞方面，“產生”的搭配，抽象義多些，如“靈感”“摩擦”，而“發生”側重具體的事情如“戰事”“疾病”等。另外值得注意的是，在左搭配中，“產生”還和“由…~/經…~/在…心裡~”“賴以~/相結合~”“之所以~”“使（人）…~”組合成典型搭配。觀察這幾組搭配組合，可以發現這些詞語搭配強調的還是“產生”的意義的外延。“產生”

表 9 本族語語料庫中“產生”和“發生”的典型搭配

	“產生”左搭配	“產生”右搭配	“發生”左搭配	“發生”右搭配
名詞 / 名詞性方位短語 +~	【主調】		【主調】	
~+ 名詞 / 動詞	靈感~/ 摩擦~/ 文字~		戰事~/ 現象~/ 面貌~/ 疾病~/ 眼前~	
		【動賓】		【動賓】
		~效應/ 疑問/ 影響/ 歧義/ 動搖/ 意義/ 後果/ 熱量/ 抗藥性/ 懷疑/ 興趣/ 飛躍/ 想法/ 幻覺/ 後代/ n+ 感/ 效果		~變革/ 變化/ 短路/ 性關係/ 共鳴/ 虧損/ 法律效力/ 變故/ 口角/ 故障/ 反應/ 衝突/ 形變/ 爭議/ 變異/ 動
副詞 +~	【狀中】		【狀中】	亂/ 火災/ 興趣/ 糾紛
	憑空~/ 一經~/ 容易~/ 很難~		新近~/ 將要~/ 極易~/ 偶然~/ 剛才~/ 突然~/ 究竟~/ 容易~/ 經常~	
詞				
助動詞 +~	【狀中】		【狀中】	
匯	會~/ 能~		會~/ 可~	
介詞短語 +~	【狀中】			
搭	由…~/ 經…~/ 在…心裡~			
動詞 / 動詞短語 +~	【狀中】			
配	賴以~/ 相結合~			
連詞 +~	【狀中】			
	之所以~			
動詞 +~	【連調】			
	使(人) …~			
~+ 趨向動詞 / 助詞		【動補】		【動補】
		~出/~了		~了/ 過+ 數詞+ 次
~+ 助詞(的) + 名詞		【定中】		
		~(的) 根源		

與“發生”的共同意義有: 表示沒有的事情出現了。但在外延上, “產生”更強調從已有中產生出新事物。因此對比分析節點詞“發生”, 可以發現“發生”並無此類典型搭配組合。

右搭配方面, “產生”和“發生”主要是搭配名詞或動詞, 形成動賓結構。這類搭配從表 5、6、7、8, 即現漢語料庫中提取的典型搭配表中, 可以看出, 所占的比例最高。但同時觀察各自的典型搭配詞語, 可以發現, “產生”的搭配詞, 在語義傾向上, 更偏向中性, 典型搭配詞有“~ 效應/ 疑問/ 影響/ 歧義/ 動搖/ 意義/ 後果/ 熱量…”; “發生”的搭配詞, 在語義傾向上, 更偏向消極, 負面, 典型搭配詞有“~ 虧損/ 變故/ 口角/ 故障/ 衝

突/ 爭議/ 動亂/ 火災/ 糾紛…”。

對比兩部學習型詞典, 可以發現兩部詞典在搭配詞語方面有重合, 但是未充分體現本族語語料庫的搭配狀況。而且從學習者的產出偏誤來看, 詞典雖然對“產生”和“發生”這組近義詞進行了辨析, 但是相關的例證, 並沒有體現出節點詞基於真實語料的特徵, 因此學習者在相關的產出上, 偏誤較多, 搭配用法也較單一。值得一提的是, 有相當比例的從本族語語料庫中抽取的高頻搭配, 並未在學習者語料中體現出來。其中的原因可能是因為學習者在產出時, 因不熟悉相關的搭配組合而傾向避免使用。這一點, 可以通過在詞典中, 增加典型搭配例證解決。

3.4 學習型詞典在例證收錄和編排方面的問題和解決方式

語料庫的大量真實語料，反映的是語言在實際生活中的真實用法。語料庫語言學的發展，同時也為詞典編撰帶來了變革。自 1987 年，John Sinclair 主編的第一部基於語料庫資料，為學習者學習現代英語而編寫的詞典—《柯林斯 COBUILD 英語詞典》誕生起，語料庫就開始為詞典的編撰提供強大的支撐。學習型漢語詞典，服務對象為學習漢語的非本族語學習者，因此在歸納漢語本族語者語言典型用法的同時，還應考慮學習者獨特的需求和因素。

3.4.1 例證突出典型搭配詞語，保證數量

例證是漢語學習者掌握易混淆詞的一個重要途徑和手段。通過對實際語言用例的學習，學習者才能掌握詞語的使用方法。作為詞語搭配的基本屬性，典型性是基於詞語的搭配力來說的。詞項與詞項之間的相互吸引相互結合，範圍有限。對吸引和結合力度最強的詞語搭配的提取，就能形成典型性搭配。學習者對典型性搭配的掌握，能夠使他們語言使用在其真實度上更接近目的語言。

《商務館學漢語》和《現代漢語學習詞典》收錄和編排的例證，部分反映了詞語的典型搭配。從上文的對比分析可以看出，以“產生”和“發生”為考察對象的例證中，左搭配和右搭配的典型性都有待提升。例如，基於語料庫提取出來的常用搭配“產生……感”“能產生”“發生了/過數詞+次”“經常發生”等，許多典型搭配組合並未呈現出來。當然因為詞典篇幅的原因，對於例證的收錄，也需要有所取捨。例證的數量也是值得探討的問題。對於近義詞來說，學習者容易在許多搭配組合上出現誤用。因此，除了採取例句的方式外，也可以採用短語例證，將更多的典型搭配組合列入，保證例證的數量，從而幫助提高學習者的詞語搭配能力。對例證的數量的強調，也並

不是說，例證越多越好，而是從詞語本身出發，既要考慮本族語的典型用法，也要關注學習者因素。不同於母語者，二語學習者對詞典的依賴性更強。這一點也可以從漢語學習者語料庫中的語言產出有所窺探，學習者所用的與“產生”“發生”相關的搭配，主要集中在詞典例證列出的其中幾個搭配組合。因此，對恰當數量的典型例證的收錄和編排，是幫助學習者學習地道語言的手段。

3.4.2 擴大近義詞辨析的範圍

本文選取的節點詞是近義詞，通過學習者語料庫的考察分析，發現在其誤用上，排在前列的除了“產生”和“發生”的誤用，還有“產生”和“生產”的誤用。對於母語者來說，“產生”和“生產”極易辨別，一般不會引起混淆。但是非本族語學習者是一個特殊的群體，這種非典型的近義詞，也極易引起辨析上的困難。因此在詞典收錄和編排時，應該更多地從學習者的角度出發，在瞭解他們相關詞語的習得難點基礎上，進行說明和配例。

3.4.3 重視近義詞搭配異同，聚焦搭配混淆點

區分近義詞詞義離不開對搭配詞的研究。從《商務館學漢語》和《現代漢語學習詞典》來看，傳統的近義詞辨析也正是這樣做的。《商務館學漢語》在“發生”和“產生”的辨析中，指出兩個詞都可以搭配“興趣”和“矛盾”，並同時給出示例。但是參照詞典的示例，有可能帶來“興趣”和“矛盾”在搭配上的偏誤。因此，就可以互用的搭配而言，也還需要考慮語言使用的語境，在有的語境中，用一種搭配可能就會比另一種搭配更為恰當。詞典既要解釋有時可以互用的語境，也要解釋有時不可以互用的語境，並給出相關的例證。另外就搭配詞語來說，除了可以互用的典型搭配詞外，還存在不可互用的典型搭配詞。也就是說，詞典在例證時，要關注搭配詞的異同，

幫助學習者更好地區分近義詞詞義。

同時需要指出的，二語學習者對近義詞的混淆點，跟母語者相比，可能存在差異。如本文提到的學習者對“產生”和“生產”的混淆。二語學習者學習近義詞時，由於混淆點不同、導致錯誤的原因各異，不宜強求詞典採取統一的辨析格式^[27]。也就是說，聚焦學習者近義詞學習的混淆點，要求詞典因類制宜，擴大漢語學習的近義詞的範圍，為不同類別的易混淆詞收錄和編排例證。

4 結語

學習型詞典的例證是學習者學習目標語的重要參考。本文在綜合考察了本族語語料庫和

學習者語料庫的基礎上，對學習型詞典的例證進行了研究。通過提取本族語語料庫中的典型搭配詞和分析學習者語料庫中目標詞語的偏誤用法，發現“產生”和“發生”這組詞，在學習型詞典中的搭配，可以從“搭配典型性”“增加短語示例”“擴大辨析範圍”以及“聚焦學習者易混淆點”等方面進行改善。學習者詞彙習得受到多方因素的影響，而學習詞典作為重要的參考工具書，具有示範和指導作用。雖然學習型詞典在編寫時無法展現本族語詞語搭配的全貌，但是語料庫驅動的研究方法，可以發揮有力的支撐功能，更大程度地考慮漢語學習者因素。

參考文獻

- [1] 劉叔新. 漢語描寫詞彙學. 北京: 北京商務印書館, 2005: 306.
- [2] Hartmann, R.R.K., & G. James. Dictionary of Lexicography. Beijing: Foreign Language Teaching and Research Press, 2002: 82.
- [3] Biber, D., Conrad, S., & Reppen, R. Corpus Linguistics Investigating Language Structure and Use. Cambridge: Cambridge University Press, 2000.
- [4] Kennedy, G. An Introduction to Corpus Linguistics. London and New York: Longman, 1998.
- [5] Thomas J., & Short M. Using Corpora for Language Research. Beijing: Foreign Language Teaching and Research Press, 2001.
- [6] 王春豔. 基於語料庫的中國學習者英語近義詞區分探討. 外語與外語教學, 2009, 6: 27-31.
- [7] 陸軍. 基於語料庫的學習者英語近義詞搭配行為與語義韻研究. 現代外語, 2010, 33(3): 276-286.
- [8] Firth, J. R. Papers in Linguistics 1934-1951. London: Oxford University Press, 1957.
- [9] Sinclair, J. The Phrase, The Whole Phrase and Nothing but the Phrase. Sylviane Granger and Fanny Meunier. Phraseology: An Interdisciplinary Perspective. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2008: 407-410.
- [10] 陸軍, 衛乃興. 短語學視角下的二語詞語知識研究. 外語教學與研究, 2014, 46(6): 865-878.
- [11] 陸方皓. 基於語料庫的對外漢語近義詞教學. 雲南師範大學學報(對外漢語教學與研究版), 2016, 14(5): 49-56.
- [12] 曾利, 陳叢梅. 漢英學習型詞典中動詞“提升”的類聯接資訊表徵研究. 辭書研究, 2020, 2: 42-51.
- [13] 曹蓉. 語料庫驅動下現代漢語“一起”的義項分立研究. 解放軍外國語學院學報, 2019, 42(1): 92-100.
- [14] 鄭定歐. 對外漢語學習詞典學芻議. 世界漢語教學, 2004, 4: 85-94.
- [15] 魯健驥, 呂文華. 編寫對外漢語單語學習詞典的嘗試與思考——《商務館學漢語詞典》編後. 世界漢語教學, 2006, 1: 59-69.
- [16] 劉曉梅, 梁青. 淺析《商務館學漢語詞典》例證的文化傳播功能. 辭書研究, 2011, 4: 109-120.
- [17] 張宏. 外向型學習詞典《漢英雙解詞典》動詞例證調查與分析. 外語研究, 2007, 4: 94-100.
- [18] 劉慧. 利用語義韻研究改進對外漢語學習詞典釋義及用例——以“可想而知”為例. 辭書研究, 2013, 6: 27-33.
- [19] Sinclair, J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
- [20] 衛乃興. 語義韻研究的一般方法. 外語教學與研究, 2002, 34(4): 300-307.
- [21] 周小兵. 對外漢語學習詞典的編寫. 辭書研究, 1997, 2: 82-86.
- [22] 章宜華. 留學生對漢語學習詞典釋義方法和表述形式的需求之探討. 華文教學與研究, 2010, 3: 50-56.
- [23] 劉川平. 對外漢語學習詞典用例的一般原則. 辭書研究, 2006, 4: 99-106.
- [24] Sinclair, J. Beginning the study of lexis. C. E. Bazell, et al. In Memory of J.R. Firth. London: Longman, 1966: 410-430.
- [25] 衛乃興. 基於語料庫和語料庫驅動的詞語搭配研究. 當代語言學, 2002, 4(2): 102-114.
- [26] 孫宏林. 詞語搭配在文本中的分佈特徵. 中文資訊處理國際會議論文集. 北京: 清華大學出版社, 1998: 230-236.
- [27] 張博. 外向型混淆詞辨析詞典的編撰原則與體例設想. 漢語學習, 2008, 1: 85-92.