

兩岸中文語料庫的構式成語偵測 與發展之異同：以 [一 A # B] 為例

施孟賢

國立聯合大學華語文學系

摘要

中文特有的四字格成語源遠流長，多從古漢語精煉而來，常具有特定的構式，且該結構能用於生成更多的新成語。兩岸的中文經過數十年的發展，已具有個別獨特的詞彙，成語的使用亦有所不同。本文運用構式搭配分析法 (Collostructional Analysis)，從繁簡中文新聞語料庫中擷取了 238 個符合 [一 A # B] 構式的繁體中文成語，以及 338 個簡體中文成語。雖然本文採用簡體中文語料庫的新聞總篇數 (1990~2002 年共 734,741 篇) 僅不到繁體中文新聞 (1991~2002 年共 1,602,174 篇) 之半數，但因大陸簡體中文的使用者較多、以及兩岸社經文化之差異，所以 [一 A 千 B] 已於大陸 (簡體中文) 發展出「一語千金／鈞」等構式成語，繁體中文則未見使用。簡體中文新發展的構式成語能否持續地使用並留存下來，繁體中文是否因使用人口較少，尚未發展出新用法，或者受語言接觸影響而傳入新用法，是未來值得觀察的方向，從構式觀點分析成語及其能產性亦有助於外語學習者理解和運用成語。

關鍵詞：構式語法、四字格成語、搭配詞、語料庫

1. 引言

中文特有的四字格成語（Quadri-syllabic Idiomatic Expressions）源遠流長，多從古漢語精煉而來（周薦，1994）。劉叔新（1990：124–127）認為成語是固定語中的一類，與其他類的慣用語和歇後語有所區別。慣用語不具意義的雙層性，而「意義的雙層性是漢語成語的區別性特徵，據此可把成語同其他固定語單位區別開，特別是同慣用語區分開，從而能以具有明確的個性而確定下來」。在各類固定語當中，劉叔新（1990：136–137）將固定語分成以下六類。A類固定語因不具雙層意義，所以屬於固定語中的慣用語一類，而與成語不同：

- A. 唯命是從、互相依存、變化無常、進退兩難
 虛張聲勢、少見多怪、不自量力、直截了當
 裝模作樣、自視過高、不堪一擊、裝糊塗
 地廣人稀、富甲天下、江山如畫、文明古國、卷帙浩繁
 量力而為、安之若素、自給自足、在所難免
 痛哭流涕、怨天尤人、喜不自勝、哈哈大笑
 不屑一顧、不言自明、不置可否、不勝其煩

B類的固定語皆具有意義的雙層性，因此屬於成語：

- B. 馬首是瞻、唇亡齒寒、風雲變幻、騎虎難下
 驢蒙虎皮、蜀犬吠日、蚍蜉撼樹、開門見山

C類的固定語其前半部分雖具有意義的雙層性，但真實的意義在停歇之後又以註釋的方式顯示出來，因此屬於固定語中的歇後語一類，須與成語區分開來：

- C. 水仙花不開——裝蒜
 水豆腐——不經打
 癩蛤蟆跳上戥盤——不知自己分量
 老雞屁股上插孔雀毛——充洋像¹

¹ 引文多為大陸用法，「充洋像」指裝模作樣之義。

D 類為專名語，指涉個別事物的概念：

D. 柴達木盆地、珠江三角洲、三潭印月、中華人民共和國、四庫全書

E 類固定語則為專門用語，特指科學技術或職業的專門概念：

E. 擴大再生產、針刺麻醉、光合作用、地心引力

F 類並未完全固定，因此部分成分可由其他字替換：

F. 大慟、不恚、大悅、一笑、一瞥、不消說、不一定、不堪

換而言之，A 類和 B 類同為四字格，但是因為後者具有表義雙層性，而被劉叔新視為成語。

另一方面，周薦（1997：29）則認為成語屬於熟語中的雅言，另一類熟語則為俗語。由於學界對於成語的定義尚未一致，他提出成語需具有「經典性」的看法，以將成語與其他熟語區別（周薦，1994：33）：

所謂經典性，是說某個熟語單位出自權威性的著作，由於該熟語所從出的著作具有權威性，熟語本身也具有了一種權威性。有些熟語源出於權威性著作，有些熟語不源出於權威性著作而出自普通的作品或乾脆就是一種口頭創作，由此顯示出兩者的區別：前者古樸、凝重、具有經典性，後者通俗、直白、不具經典性。

成語多出自權威性的著作，如十三經、官修和私撰的正史、子書和集書中的名家名作，因而具有經典性。成語之外的他類熟語，如諺語、歇後語、慣用語，則很少出自權威性的著作，多是隨機的口頭創作或出自俗白的作品，因而不具經典性（周薦，1997：29）。

根據中國大陸《漢語新成語詞典》，其收錄的新成語則不僅止於曾出現於經典古文的四字格，此新成語詞典所收錄的成語共有七類（史式、趙培玉，2002）：

1. 漢語普通話區和各方言區口頭上的熟語與諺語，五四運動以前書面上不用或者少用，以後才在現代作品中廣泛使用的。
2. 外來成語，包括外國的與國內少數民族的成語，五四運動以後通過翻譯進入漢語的。
3. 文言中的老成語，在五四運動以後又產生了新義，現在廣泛使用的。
4. 古典詩詞原句，五四運動以後用為成語。
5. 五四運動以後現代作者創造的新成語。
6. 五四運動以後，各門科學術語經過群眾廣泛使用而形成成語的。
7. 其他為表達一種新的概念而經群眾創造出來的新成語。

值得注意的是，《漢語新成語詞典》將外來語和新成語也納入新成語，此一原則可做為未來新形成的四字格是否納入成語的參考。

學界另有認為成語具書面語性質的看法，劉潔修（1985：23）提出「書面語為成語，口頭語為俗語」。武占坤（1983：104）認為「成語語體風格，屬於書面語詞彙的性質……從而和諺語、歇後語、慣用語等區別開來」。馬國凡（1983：64）指出「成語所用的語言成分，帶有強烈的書面語性質」。劉德馨等人（2019）則從「詞素黏著度」切入，凸顯四字格成語的獨特性：四字格的詞素黏著度高，無法加入其他的文字，句式也不可任意更動，例如「一壺千金」不能改成「*一壺需要千金」。三字格慣用語則不然，「開夜車」可以改成「開了一整晚的夜車」。值得注意的是，四字格成語雖屬「固定詞組」，但隨著時代演進、政策需求，新造的四字格不斷出現，例如「一例一休」和「一帶一路」等，顯示四字格成語具有相當程度的規律性和能產性。由於中文成語具有上述的表義雙層性、經典性和書面語性質，對華語學習者而言，更具挑戰性。而成語的數量龐大，如何以有效率的方式提升學習成效，亦是華語教師所面臨的難題。

從「一例一休」和「一帶一路」等新創四字格，可以看出成語常具有特定的構式，且該結構能用於生成更多的新成語。根據 Goldberg（1995: 5）對構式的定義：

C is a CONSTRUCTION iff_{def} C is a form-meaning pair $\langle F_i, S_i \rangle$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions.

也就是說，當 C 的意義無法從其本身的組成成分或其他構式推知時，C 即為一個構式。由於多數成語內部皆具有構式之存在，其意義無法從個別詞素的相加推導得出，語義透明度低，遂對不熟悉構式義的外籍學習者造成困擾。

Su (2002: 31) 首先將構式語法用於分析中文的四字格，她認為四字格中的 [X 來 Y 去] 構式帶有重複的涵義，例如「翻來覆去」、「走來走去」。劉德馨等人 (2019) 進一步討論 [一 X # Y] (「#」表數字) 構式，並分析三個類型頻率 (type frequency) 最高的具體形式 [一 X 一 Y]、[一 X 二 Y] 和 [一 X 千 Y]，著手探討構式成語的語法功能。該文發現四字格成語具有特殊的構式意義，一個成語的意義來自於 XY 名詞的多義性和成語構式之間的互動關係，兩者的互動構成了中文四字格成語的規律性和能產性。此外，嵌入四字格成語的雙字詞 XY 多為近義詞或反義詞，例如「一心一意」、「一生一世」之嵌入成分為近義詞，「一來一往」、「一薰一蕕」之嵌入成分為反義詞，「一例一休」和「一帶一路」兩個新創四字格，其嵌入成分也皆為近義詞。而為了幫助學習者掌握構式成語的規律，劉德馨、施孟賢 (2022) 分析了 1095 筆中文四字格構式成語，依構式義分門別類，並標記構式成語的語法功能和提供例句，製作「構式成語語料庫」²。此一構式成語語料庫和北京大學的現代漢語構式數據庫³不同，北京大學之語料庫透過人工和半自動的方式從大規模真實語料中（以北京大學 CCL 現代漢語語料庫為主）實際運用的構式逐一標註，並詳細描述每個構式內部的構成情形，整體的句法、語義及語用特性（詹衛東，2018：6）。「構式成語語料庫」則分析構式和嵌入成語的雙字詞的互動關係，以及構式成語於語法、語義和語用之規律性，方便教師將相關例句應用於華語教學和教材編寫。

中文成語的結構具有規律性，根據 Tsou (2012: 51) 的統計，對稱結構 (Coordinative) 的成語如「任重道遠」，佔所有成語的 35%。若能從構式的角度切入，系統化地帶入一系列結構類似的成語，搭配其構式義進行教學，相信學習者便能快速掌握大量的成語，克服表義雙層性的困境。

立基於上述研究，本文試圖運用構式搭配分析法 (Collostructional Analysis, Stefanowitsch & Gries, 2003)，從兩岸繁簡中文大型語料庫中擷取構式成語，進而探討構式成語的差異，並聚焦於成語構式 [一 A # B] 於兩岸中文成語之能產性。

² <http://idiomconstructions.github.io>

³ <http://ccl.pku.edu.cn/ccgd>

2. 構式搭配強度計算

四字格搭配強度的估算，可從傳統搭配詞強度 MI (Mutual Information) 的計算進行延伸。MI 值指的是兩詞實際觀察到共同出現的機率 (observed probability) 與兩獨立詞分別出現的機率 (expected probability) 之間的比例：

$$(1) MI(w_1, w_2) = \frac{O(w_1, w_2)}{E(w_1, w_2)}$$

但古典的 MI 值僅能用於計算兩者之間的搭配強度，不適用於三者以上(如四字詞)的搭配強度估算。da Silva & Lopes (1999: 372) 將原本的 MI 值推展至三者以上的運算，並命名為 G 值 (Glue value)：

$$(2) G(w_1, w_2, w_3) = \frac{p(w_1, w_2, w_3)}{p(w_1, w_2) \cdot p(w_3)}$$

然而 G 值的運算須將語料庫中每個詞的單獨詞頻和各種可能的搭配詞組合的頻率皆計算過，才能得出最高的搭配組合。對大型語料庫而言，需耗費大量運算時間和資源，因此本文採用 Stefanowitsch & Gries (2003) 的構式搭配分析法 (Collostructional Analysis)。該分析法用於計算構式 (construction) 與詞位 (lexeme) 之間的強度，採用費雪精確檢定 (Fisher exact test)，如公式 (3)，計算出目標詞位與特定構式之間的搭配強度。精確檢定無須假設樣本分布，因此適用於小樣本 (Pedersen, 1996)。由於語言樣本中少數高頻詞占最大比例，而大多數詞的出現頻率相對而言皆很低 (Zipf, 1935)，因此比起 Person's χ^2 、Likelihood ratio G^2 以及 t-test 可能有違反大樣本假設的問題，Fisher exact test 較適用於語言樣本 (Pedersen, 1996)。

$$(3) p_{\text{Fisher exact}}(a, b, c, d) = \frac{C_a^{a+c} C_b^{b+d}}{C_{a+b}^{a+b+c+d}}$$

以英文詞位 *accident* 出現在構式 [N waiting to happen] 中的搭配強度計算為例 (Stefanowitsch & Gries, 2003: 219)，首先以列聯表 (cross tabulation) 表 1 檢視各類搭配組合的出現次數：

表 1 英文詞位 *accident* 與構式 [N waiting to happen] 的次數列聯表
(Stefanowitsch & Gries, 2003: 219)

	<i>accident</i>	\neg <i>accident</i>
[N waiting to happen]	a = 14	b = 21
\neg [N waiting to happen]	c = 8,606	d = 10,197,659

列聯表中的 a、b、c、d 分別是目標詞位 *accident* 在語料庫（本例為 British National Corpus 英國國家語料庫 BNC）與構式 [N waiting to happen] 一同出現的次數，b 是其他非 *accident* 的詞位（例如 *disaster*, *recovery*, *dream*, *event* 等等）出現在構式 [N waiting to happen] 中的次數，c 是目標詞位 *accident* 出現在非 [N waiting to happen] 構式的次數，d 為其他非目標詞位未出現在 [N waiting to happen] 的次數。

從列聯表 1 中各類搭配組合 a、b、c、d 的分布情形，即可依據前述費雪精確檢定的公式（4），計算出詞位 *accident* 與構式 [N waiting to happen] 的搭配強度，如下所示：

$$(4) p_{Fisherexact}(14, 21, 8606, 10197659) = \frac{C_{14}^{8,620} C_{21}^{10,197,680}}{C_{35}^{10,206,300}} = 2.1 \times 10^{-34}$$

對中文四字格成語而言，以「千呼萬喚」為例，中文構式 [千 X 萬 Y] 即相當於上述的英文構式 [N waiting to happen]，而本文將 XY 視為一詞位，相當於上例的 *accident* 詞位，並且在以下的列聯表 2 呈現 [千 X 萬 Y] 各搭配組合次數：

表 2 構式 [千 X 萬 Y] 與「呼喚」詞位的搭配次數列聯表

	呼喚	\neg 呼喚
[千 X 萬 Y]	a = 103	b = 16,658
\neg [千 X 萬 Y]	c = 403	d = 1,716,668

我們同樣可用費雪精確檢定，計算構式 [千 X 萬 Y] 與「呼喚」詞位的搭配強度如（5）所示：

$$(5) p_{Fisherexact}(103, 16658, 403, 1716668) = \frac{C_{103}^{506} C_{16,658}^{1,733,326}}{C_{16,761}^{1,733,832}} \\ = 2.29 \times 10^{-100}$$

經過費雪精確檢定，p-value 的結果為 2.29×10^{-100} ，表示構式 [千 X 萬 Y] 與「呼喚」詞位的搭配強度為顯著。藉由計算構式 [千 X 萬 Y] 與語料庫中其他各種搭配詞位之列聯表次數，我們可以比較出與構式 [千 X 萬 Y] 搭配強度較高之詞位，進一步找出可能的構式成語。

立基於以上針對中文構式成語的搭配分析法，本文聚焦於以下三個研究問題：

1. 構式搭配分析法 (collostructional analysis) 能否用於大型語料庫中構式成語之偵測 (以費雪精確檢定為例)？其成效如何？
2. 兩岸繁簡中文的構式成語發展有何異同？
3. 新創之四字格構式，其組成規則是否符合構式成語之構成條件？

本文探討四字格成語中的構式 (如「一呼百應」中的 [一 A 百 B])，並將嵌入成語構式中之 A、B 視為前述構式搭配分析法中的詞位，進而將成語構式在語料庫中與各種詞位的搭配強度進行排序，嘗試利用費雪精確檢定擷取兩岸繁簡中文語料庫中的構式成語，並比較兩岸構式成語發展之異同。

3. 研究方法

本文的語料來自兩岸繁簡中文同時期的新聞，如表 3 所示，繁體中文語料則來自臺灣中央通訊社從 1991 年至 2002 年，共 1,602,174 則新聞 (約 7.3 億字)；簡體中文的新聞來自大陸新華社自 1990 年至 2002 年的 734,741 則新聞 (約 3.8 億字)。

表 3 兩岸繁簡中文新聞語料庫統計資訊

新聞來源	年代	文章數	字數
中央通訊社	1991 - 2002	1,602,174	7.3 億繁體字
新華社	1990 - 2002	734,741	3.8 億簡體字

大型中文語料庫因人力成本關係，無法逐一人工檢驗自動斷詞或詞類標記的結果，因此容易出現斷詞錯誤的情形。本文研究目標為四字格構式成語，一旦發生斷詞錯誤，則後續之四字成語皆無法正確擷取，故本研究所有的中文語料皆未經過斷詞或詞類標記，而直接以原始新聞語料，計算其中目標構式在語料庫中與其他嵌入字詞之頻率，進而求得其搭配強度。

以 [-A 千 B] 為例，與該構式搭配強度最高的四字格為「一落千丈」，如表 4 所列，其費雪精確檢定 p 值幾乎趨近於零（p 值越小，其搭配強度越高）。再者為新聞語料庫中多次出現的電力單位「一千千瓦」、「一萬千瓦」，接著則是常見的成語「一日千里」、「一瀉千里」、「一擲千金」，之後搭配強度較弱者皆非成語。除了根據前述成語需具表義雙層性、經典性和書面性質的判斷原則進行人工判準之外，尚可針對在 [-A #B] 構式中常見的電力單位（如「一萬千瓦」），以及常用的中文分類詞撰寫程式規則進行後續過濾（如「一個千萬」）。

表 4 與 [-A 千 B] 搭配強度最高的四字格

Collexeme	$P_{\text{Fisher_exact}}$ (collostruction strength)
一落千丈	0.00
一千千瓦	3.02×10^{-276}
一萬千瓦	4.11×10^{-199}
一日千里	7.86×10^{-170}
一瀉千里	2.51×10^{-85}
一擲千金	1.97×10^{-66}
一百千瓦	2.23×10^{-35}
一個千萬	8.35×10^{-33}
一個千載	5.70×10^{-25}

4. 結果與討論

本文探討構式成語 [一 A # B] 於兩岸中文發展之差異，並聚焦於 [一 A 二 B]、[一 A 百 B]、[一 A 千 B] 三個構式，經過人工檢核後，總共於繁體中文的新聞語料庫中擷取出 238 個符合 [一 A # B] 構式的成語，簡體中文的語料中則有 338 個構式成語。表 5 列出個別 [一 A # B] 四字格於語料庫中出現的頻率，各個構式和詞位的頻率以及搭配強度的細節請見附錄。以下分別呈現三個構式的結果以及討論：

表 5 [一 A # B] 四字格於繁簡中文語料庫中出現的頻率統計

	繁體中文語料的頻率	簡體中文語料的頻率
[一 A 二 B]	7,020	478
一乾二淨	117	45
[一 A 百 B]	3,061	196
一了百了	47	3
一唱百和	0	1
[一 A 千 B]	1,059	41
一刻千金	0	3
一字千金	6	6
一言千金	0	1

在繁體中文新聞和簡體中文新聞語料庫中，「一字千金」出現的次數皆為 6 次，但繁體中文語料庫的規模（文章篇數和字數）幾乎為簡體中文的兩倍，「一字千金」在繁體中文的使用頻率相對不高，由此可推測何以「一字千金」並未出現在表 4 中，其與 [一 A 千 B] 構式的搭配強度反而不及「一百千瓦」等非成語形式。雖然可以考慮純粹以所有四字格的出現頻率做為成語的門檻，然此做法可能另外衍生中文常用詞串（Lexical Bundles，參考 Hsu, 2021; Hsu et al., 2017; 劉貞好等人，2017）的問題。例如前述表 4 中的「一個千載」，可能因為常用詞串「一個千載難逢的機會」的關係，使得「一個千載」在語料庫中被認定為高頻的四字格，但這並非本文所討論的成語。因此，成語在語料庫的地位為何？我們是否應將成語視為一不可切割的詞彙單位，還是可切分出有內部結構的構式？或者應從不定長度的常用詞串角度處理語料庫中的成語？此課題在語料庫語言學中仍未有定論。

4.1 [-A 二 B]

在未經斷詞處理的簡體中文新聞語料庫中，[-A 二 B] 總共出現 478 次，包含「一干二淨」45 次，如例（6）；以及「一光二淨」1 次，如例（7）。

（6）临走时他把宿舍打扫得一干二净，迎接新所长的到来。

（7）长哨营乡 57 岁的炊事员、共产党员王恩普家里的东西被冲得一光二净，爱人几次找他回家看看，他却说...

繁體中文亦有「一乾二淨」，於語料庫中出現 117 次，如例（8），但未出現類似簡體中文「一光二淨」的用法。

（8）...到金門海域作業，以拖網方式將海底的生物刮得一乾二淨，滿潮時炸魚，退潮時電魚，還破壞蚵椿...

值得注意的是，大陸新創四字格「一光二淨」之嵌入成分「光」、「淨」也隸屬同一個語義場，不脫離構式成語之組構原則。

4.2 [-A 百 B]

兩岸的中文新聞語料庫中，雖然皆有「一呼百應」，如（9），和「一呼百諾」的使用，但大陸的簡體中文還發展出「一通百通」和「一順百順」的構式成語，如例（10）和（11），此兩筆成語均未見於繁體中文的語料庫。

（9）好几次，乡农技员知道后及时与农民们商量，结果一呼百应，马上给姚海根送来虫源。

（10）掌握了解放思想、实事求是这个精髓，我们就可以一通百通，做工作就可以由必然王国逐步进入自由王国。

（11）坚定地、全面地、创造性地坚持建设有中国特色社会主义理论，顺应历史，顺从民心，顺理成章，一顺百顺，我们的国家就会强盛，人民就会幸福，社会就会长治久安。

4.3 [一 A 千 B]

對 [一 A 千 B] 構式而言，兩岸中文語料庫中皆有典型的成語，如「一諾千金」和「一擲千金」，惟簡體中文的新聞語料中發展出類似的「一言千金」、「一语千金／鈞」、「一擲万金」之使用。

- (12) 当许多政府计划因为缺少资金而迟迟无法实行的时候，大企业却在篮球赛事上一掷万金。商业化的体育赛场一定要用黄金铺地吗？
- (13) 为了实践清华人一言千金的品行，清华校友团纷纷解囊相助，竟然筹资 3000 万元在清华建造了一座 3000 平方米的赛艇馆，在国际上都属首屈一指，令人咋舌。
- (14) 南钢党委书记肖同友感慨万分：黄教授一语千金，却不取分文。
- (15) 江泽民总书记的论述，精辟透彻，一语千钧。

其中「一言千金」和「一语千金」的語義與「一諾千金」相同，皆表示信守承諾，「一语千鈞」則更強調講話的份量，脫離了「一諾千金」的典故。根據劉德馨等人 (2019) 針對「一 X 千 Y」的分析，此一預製模組 (prefab) 之構式義為「對比」，例如「一日千里」(很短的時間內便跑了千里)、「一字千金」(一個字價值千金)。觀察大陸新造成語「一语千鈞」，「鈞」為古代計算重量的單位，三十斤為一鈞，因此「一语千鈞」的構式義也是「對比」，意指一句話價值千斤重，非常有份量。換而言之，新造之四字格仍不脫離 [一 A 千 B] 之原有構式義。

5. 結論

本文於兩岸繁簡中文大型語料庫中，運用構式搭配分析法 (費雪精確檢定)，分別擷取出符合 [一 A # B] 構式的 238 個繁體中文成語，以及 338 個簡體中文成語。雖然搭配強度經排序過後的四字格仍需人工檢核是否含有非成語的四字格，但對大型語料庫中的四字格成語而言，相較於單採頻率統計或視為中文常用詞串 (Lexical Bundles) 的做法，構式搭配分析還是比較適合用於四字格構式成語的方法。透過本文探討構式成語於語料庫中的搭配分析，希望對成語甚至是構式在語料庫中的定位有所貢獻。

本文為首先針對大型未斷詞語料庫中，構式搭配分析應用於中文四字格成語之研究，未來可與四字格的頻率統計、以及常用詞串的方法進行比較，以回應成語與構式於語料庫中的標記問題。

比較繁簡體語料庫後，本文發現在 [一 X 百 Y] 的構式分析中，兩岸中文皆有「一呼百應」和「一呼百諾」的使用，不過簡體中文還另外發展出新的「一通百通」、「一順百順」用法；兩岸中文的 [一 X 二 Y] 構式皆能形成「一乾二淨」等成語，簡體中文另以該構式生成「一光二淨」成語；兩岸中文的語料庫中皆有經典的「一擲千金」和「一諾千金」成語的使用，但簡體中文開始發展出「一擲萬金」、「一語千金」與「一言千金」的類似用法。本文發現新創之四字格，其構詞規則仍然符合 [一 X # Y] 之組構原則，也就是說，屬於固定詞組的成語，同時具有某種程度的開放性和能產性（劉德馨等人，2019）。

雖然本文採用簡體中文語料庫的新聞總篇數（1990~2002 年共 734,741 篇）不到繁體中文新聞（1991~2002 年共 1,602,174 篇）之半數，然推測因同一時期大陸簡體中文的使用者較多，以及兩岸社經文化之差異，遂以 [一 X 千 Y] 發展出「一語千金／鈞」等未見於繁體中文語料庫中的構式成語。

簡體中文新發展的構式成語能否持續地使用並留存下來，繁體中文是否因使用人口較少，尚未發展出新用法，或者是否受語言接觸影響而傳入新用法，是未來值得觀察的方向。此外，從構式觀點分析成語及其能產性亦有助於外語學習者理解和運用成語。

謝辭

本研究感謝兩位匿名審稿委員、主編以及校稿者之意見，文中若尚有未善之處，為本文作者之責任。

附錄

[一 A#B] 構式於語料庫中與各詞位的搭配強度

	成語頻率	對應構式	構式頻率	詞位	詞位頻率	$P_{\text{Fisher exact}}$
1	一心一意	493	一.一.	心意	773	0
2	一言一行	210	一.一.	言行	1,370	4.42E-173
3	一朝一夕	166	一.一.	朝夕	460	3.65E-190
4	一草一木	139	一.一.	草木	408	1.26E-156
5	一兵一卒	80	一.一.	兵卒	5	9.69E-159
6	一板一眼	31	一.一.	板眼	7	8.96E-58
7	一唱一和	29	一.一.	唱和	224	1.69E-23
8	一生一世	19	一.一.	生世	100	1.09E-18
9	一五一十	19	一.一.	五十	27,722	5.21E-76
10	一心一德	5	一.一.	心德	260	0.07880821
11	一字一板	5	一.一.	字板	12	2.59E-07
12	一死一生	2	一.一.	死生	10	0.00455286
13	一手一足	2	一.一.	手足	448	0.60123226
14	一悲一喜	1	一.一.	悲喜	150	1
15	一厘一毫	1	一.一.	厘毫	1	0.0170182
16	一目十行	1	一.十.	目行	45	0.06163976
17	一波三折	239	一.三.	波折	420	0
18	一日千里	94	一.千.	日里	970	2.13E-192
19	一字千金	6	一.千.	字金	9	7.26E-18
20	一刻千金	3	一.千.	刻金	2	3.83E-10
21	一言千金	1	一.千.	言金	3	0.00134893
22	一清二楚	192	一.二.	清楚	7,898	1.20E-86
23	一身二任	9	一.二.	身任	26	9.92E-15
24	一清二白	4	一.二.	清白	561	0.16674831
25	一林二虎	1	一.二.	林虎	146	0.42607551
26	一文不值	21	一.不.	文值	1	1.32E-45
27	一字不差	16	一.不.	字差	22	1.19E-25
28	一病不起	11	一.不.	病起	42	3.69E-14
29	一文不名	5	一.不.	文名	233	0.01815983
30	一事不知	2	一.不.	事知	100	0.13569155
31	一言不合	1	一.不.	言合	13	0.08466255
32	一知半解	36	一.半.	知解	10	1.17E-91
33	一男半女	1	一.半.	男女	10,726	8.70E-07
34	一了百了	3	一.百.	了了	602	0.02206196
35	一唱百和	1	一.百.	唱和	224	0.1971167

參考文獻

- 劉叔新 (1990)。漢語描寫詞彙學。北京：商務印書館。
- 劉德馨、呂佳蓉、蔡宜妮、蘇以文 (2019)。四字格成語的習語性和規律性——以「一X#Y」為例。清華學報，49 (4)，683—719。
- 劉德馨、施孟賢 (2022)。建置構式成語語料庫輔助華語教學——兼論同型構式和近義構式。華語文教學研究，19 (1)，95—121。
- 劉潔修 (1985)。成語。商務印書館。
- 劉貞好、陳浩然、楊惠媚 (2017)。中文人文社會科學論文常用詞串之研究。華語文教學研究，14 (1)，119—152。
- 史式、趙培玉 (2002)。漢語新成語詞典：1919-2001。重慶：重慶出版社。
- 周薦 (1994)。熟語的經典性和非經典性。語文研究，(3)，33—38。
- 周薦 (1997)。論成語的經典性。南開學報，(2)，29—35。
- 武佔坤 (1983)。詞彙。上海：上海教育出版社。
- 詹衛東 (2018)。現代漢語構式知識庫。北京：北京大學出版社。
- 馬國凡。(1983)。成語。呼和浩特市：內蒙古人民出版社。
- da Silva, J. F., & Lopes, G. P. (1999). A local maxima method and a fair dispersion normalization for extracting multi-word units from corpora. In *Proceedings of the 6th Meeting on the Mathematics of Language*, 369–381.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Hsu, C.-C. (2021). The Structure and Function of Lexical Bundles in Chinese Conversation and News. *Taiwan Journal of Chinese as a Second Language*, (22), 69—96.
- Hsu, J.-W. K., Chen, H.-J. H., Yang, T.-Y. C., & Liu, C.-Y. (2017). A study on the use of lexical bundles in native and nonnative English academic writing. *English Teaching & Learning*, 41(2), 55—98.
- Pedersen, T. (1996). Fishing for Exactness. In *Proceedings of the South-Central SAS Users Group Conference (SCSUG-96)*, 188—200.
- Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209—243.
- Su, I.-W. (2002). Why a construction-that is the question! *Concentric: Studies in*

Linguistics, 28(2), 27–42.

Tsou, B. K. (2012). Idiomaticity and classical traditions in some east Asian languages. In *Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 2012)*, 39–55.

Zipf, G. K. (1935). *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. London: Routledge.

[收稿：2024.01.02；修訂：2024.01.12；接受：2024.01.18]

施孟賢 Meng-Hsien Shih

國立聯合大學華語文學系

Department of Chinese Language and Literature

National United University

mhshih@nuu.edu.tw; simon.xian@gmail.com

TEL: 037-382704

Detecting and Analyzing the Development of Quadri-syllabic Idiomatic Constructions in Cross-strait Chinese Variation: A Case Study of [一 A # B]

Meng-Hsien Shih

National United University

Abstract

The special quadri-syllabic idiomatic expressions (QIEs) in Chinese have existed for a long history, mostly refined from archaic Chinese and consisting a specific construction which can be used to generate more new QIEs. Over half a century of development cross the strait, both Chinese variations have evolved special words and differed in the uses of certain idioms. This paper adopts Collostructional Analysis to a news corpus, extracting 238 QIEs in traditional Chinese characters with the [*one-A-number-B*] construction, and 338 simplified Chinese QIEs. Although the total number of news articles in the simplified Chinese corpus used in this paper (734,741 articles from 1990 to 2002) is less than half of those in the traditional Chinese news corpus (1,602,174 articles from 1991 to 2002), it is speculated that due to a larger user base of simplified Chinese in mainland China during the same period and differences in socio-cultural aspects between the two areas, constructions like [*one-A-thousand-B*] led to the emergence of idioms such as 一語千金 ‘one word (worth) thousand gold’, which was not found in the traditional Chinese corpus. It is worth observing in the future whether these newly developed QIEs with constructions in simplified Chinese will continue to be used and preserved, and whether traditional

Chinese, due to a smaller user base, has yet to develop new usages. It is also interesting whether further language interactions will introduce new usages in the future. Analyzing idioms and their productivity from a perspective of construction can also assist foreign language learners in understanding and using idiomatic expressions.

Keywords: construction grammar, quadri-syllabic idiomatic expressions, collocation, corpus