

基於大型語料庫的現代漢語離合詞定量研究*

王海峰

北京大學對外漢語教育學院

摘要

本文借鑒語料庫驅動研究方法（corpus-driven approach）的理念，利用北大語料庫對《漢語水準辭彙與漢字等級大綱》中 207 個離合詞進行考察，得出結論：1.所謂離合詞，實際上“合”是常態（regular），“離”是異態（idiosyncratic）；2.離合詞插入“了”的離析形式最多，其次是插入補語或插入名詞/代詞以及數量詞的離析形式，而插入動詞性形式的情況最少；3.離合詞離析形式的出現種類與以這些形式離析的離合詞的數量和所占比例成反比；4.離合詞的離析頻率與使用頻率正相關。

關鍵詞：語料庫，離合詞，定量，離析形式，頻率

1. 定量研究在離合詞教學與研究中必要性

現代漢語離合詞是當前漢語教學和研究的重點和難點，多年來人們對離合詞的性質、用法、語法規律等各方面進行了許多有益的探索。但令人遺憾的是，許多問題我們仍然沒有得出一致的結論。筆者認為主要原因之一是，我們在討論離合詞問題時忽視了語言的客觀實際，多憑語感去主觀設想，因而一些問題眾說紛紜，莫衷一是。值得特別提出的是，我們至今並沒有完全弄清離合詞這個家族中每個詞離析方式的詳細情況。連離合詞的這個最基礎的問題都沒有搞清楚，離合詞的教學與研究就只能是隔靴搔癢、盲人摸象。

* 本文的寫作得到了導師趙金銘先生的悉心指導以及詹衛東、薛晶晶等先生的幫助。文章曾在第 15 屆國際中國語言學會議暨第十九屆北美洲中國語言學研討會聯合會議（2007 年 5 月 25 日—27 日，美國紐約哥倫比亞大學）上宣讀，陸儉明、李亞非等先生的意見建議使筆者受益良多，會後筆者作了較大修改。感謝匿名審稿人提出的寶貴意見，筆者認真採納並作了一定修改。文中謬誤、不足皆由筆者本人負責。

“說到離合詞，一般的語法書都指出可以在離合詞中插入‘了、著、過’可以插入補語和定語。如果直接把這些規律搬到對外漢語教學中就顯得線條太粗、太籠統。”（饒勤 1997）饒勤的感覺是正確的。我們看離合詞“當面”：我們通過檢索北大語料庫發現其離析形式只有“著、名詞/代詞、了”3種，而且 451 條例句中，插入“著”的有 427 例，占 95%，插入“名詞/代詞、了”僅占 5%，說明只有“著”是“當面”的優先插入形式，客觀語言調查（語料庫檢索）中沒有發現有插入“過”、補語等的情況，說明不同的離合詞離析方式並不一樣。

也有很多人認識到離合詞中間插入的形式並不相同，但每個離合詞中間到底能插入什麼形式，常常憑個人語感，實際上離合詞的離析方式遠非我們想像的那樣。《現代漢語離合詞用法詞典》（1995）解釋“把關”的用法：

【離】

- （1）“把”後帶助詞“著”、“了”、“過”、“的”等：產品生產的最後階段他一直把著關，你放心好了。/你太不注重產品品質了，你什麼時候去把過關？/他們是怎麼把的關？
- （2）“把”後帶補語：產品品質方面，你必須把好關。
- （3）“關”提前到“把”前：你一定要把關把好。
- （4）重疊“把”：你們當廠長的在產品品質方面得把把關。
- （5）“關”前帶定語：你要我把什麼關？

客觀語言事實中“把關”離析情況如何呢？我們利用北大語料庫進行搜索共得出“把關”有效離析句共 361 條，分析發現“把關”的離析情況遠非如此，請看：

表 1：“把關”離析情況一覽表

離析形式 數 值	I (把+名詞+ 關)	II 把+補語+ 名詞+關	III 把+數量詞+ 關	IV 名詞提前 (關...把)	V 把+了+關
數 量 共 361 條	175	177	5	2	1
比率 (100%)	48.5%	49%	1.4%	0.6%	0.3%

例如（以下各例均出自 CCL 語料庫，恕不一一注明）：

- 1) 中國長城鋁業公司中州鋁廠強化生產現場管理，「把好產品品質關」，

優質品率大大提高，贏得了顧客的青睞。(市場報 1994)

2) ……，其次是實行工廠、車間、大組、小組四級管理體制，層層「把質量關」。(人民日報 1995)

通過語料，我們看出，“把關”的優先離析形式是 I “把+名詞+關”和 II “把+補語+名詞+關”(實際上，二者可以合併為“把+名詞+關”形式)，二者相加達 97.5%。插入“了”的情況很少，插入形式“著、過、的”和“什麼”沒有出現，“把”重疊的用法未出現¹。

由此看來，解決現代漢語離合詞的教學與研究的難題首先要從摸清離合詞的“底細”做起，而這個工作僅憑語感是不準確的，必須借助於大型的語料庫進行定量研究。

2. 利用語料庫考察離合詞的基本理念及步驟

2.1 基本理念

本研究主要借鑒語料庫驅動的研究方法 (corpus-driven approach)，這種方法主張語言學家要關注語料庫本身，語言描述必須全面關照語料庫提供的證據。理論陳述應完全和語料證據相符，即理論直接從語料證據中歸納出來並反映這些證據，語料必須保持原貌，不經任何加工。出現頻率較高的型式及其分佈是語法範疇的最基本證據。

語言表達的可能性與典型性應區分開，有些表達儘管在理論上有可能，但在實際使用中很少出現或幾乎不出現；而另外一些表達上不但理論上可能，而且在實際使用中出現的頻率很高，這是典型的表達。語料庫驅動的研究方法關注的是後者。(Hunston&Francis 2000，Tognini-Bonelli 2001，王勇 2008)

2.2 選定語料庫

本文考察統計主要使用北京大學漢語語言學研究中心(CCL)語料庫(以下簡稱 CCL 語料庫，<http://ccl.pku.edu.cn>)網路版現代漢語部分，該部分共有 157 個檔，23 個檔夾，229,700,316 位元組，約 1.15 億字。CCL 語料庫是一個比較成熟的語料庫，容量大，檢索方便；語料分佈涉及大陸、臺灣、香港等華語區，語體包括新聞、文學、口語、戲劇、應用文、翻譯作品等(詳細情況請見附錄三)。

¹ 沒有出現並不說明該離合詞沒有插入上述形式的可能，只說明該離析方式的出現率很低。尤其是在教學中，不太可能出現的離析方式，可以不必作為教學重點。

2.3 初步圈定考察範圍

2.3.1 依據《漢語水準辭彙與漢字等級大綱》結合《現代漢語詞典》

《漢語水準辭彙與漢字等級大綱》(國家漢辦 1992, 以下簡稱《大綱》)對離合詞的處理方法是“對於所謂的‘離合詞’,一律算‘離’,不標注詞性,如‘建交’,‘打倒’,‘推翻’”²。《現代漢語詞典》(第5版,以下簡稱《詞典》)對離合詞的標注方法是拼音中加“//”,如:洗澡 xǐ//zǎo。這樣,我們可以依據上述二者初步確定考察物件。

不過,二者有一致的地方也有不一致的地方³,我們的做法是,只要二者有一者認定為其離合詞,我們就將其納入考察範圍。

2.3.2 儘量全面考察

還有一些《詞典》跟《大綱》都沒有認定為離合詞的雙音結構,我們語感上認為這些可能產生離析,如:報銷、出境、登記、告別、集資、及格、慷慨、幽默、有意、有名、致敬、註冊、注意等,我們也將其納入考察範圍。

以上做法主要是為了儘量取其最大值,全面考察離合詞。通過以上甄別,我們確定了《大綱》中的392個述賓⁴、述補等雙音結構作為考察物件。

2.4 考察統計分析

我們通過CCL大型語料庫對《大綱》中的離合詞逐個進行檢索,然後進行鑒別、整理、分析,以期摸清離合詞家族的底細,為對外漢語教學與研究

² 《大綱》對一些常用結構也不標注詞性。我們的做法是,為保證考察全面,先將可離析雙音結構選入,甄別整理工作待後進行。

³ 比如《現代漢語詞典》認定為離合詞,而《大綱》沒有認定其為離合詞。如:編號、登記、定價、考試、留神、留心、留意、冒險、拍照、請願、送行、隨便、隨意、談天、同學、致詞、致電、抓緊。

也有《大綱》認定其為離合詞,但《現代漢語詞典》並未將其視為離合詞。如:定點、結果、據說、據悉、廉政、聽說、與會、越冬。

還有《大綱》列為詞條並未加詞性(即《大綱》認定為離合詞),但《現代漢語詞典》中無此詞條的,如:敬酒、吸煙(抽煙)、飲水、有害、做工。

⁴ 我們對述賓結構並不嚴格界定,比如:洗澡、睡覺、幫忙、游泳、據說、吃驚等,我們均視為述賓結構。在此問題上,已有前賢從理論和實踐上嘗試。呂冀平、戴昭銘(1985)認為“‘洗澡’從前是並列關係,現在早已變為動賓關係。近年來‘考試’‘游泳’等也似乎有這種趨勢……我們如果承認‘洗澡’是動賓關係,那麼對‘游泳’大概就得一視同仁,因為現在它們已經非常接近。”陳建民《漢語口語》(1984,北京出版社)也將“洗澡”等列為動賓式(述賓式)。

服務。

3. 離合詞及其離析形式考察統計結果及分析

3.1 可離析雙音結構考察統計分析

我們通過 CCL 語料庫對 392 個可離析雙音結構逐個檢索，共得出 34，441 個可離析雙音結構的有效離析例句。對每個雙音結構的離析情況進行比對分析，我們發現每個可離析雙音結構離析形式出現的數量差異較大，最多的出現 1700 多例，還有很多則一例也沒有出現。

請看下表：

表 2 可離析雙音結構離析形式數量分佈表

例數 數值	1000 例以上	999-500 例	499-100 例	99-10 例	9-1 例	0 例
數量(共 392 個)	7	9	54	162	115	45
比率(100%)	1.8%	2.3%	13.8%	41.3%	29.3%	11.5%

統計發現，未出現離析情況的雙音結構有 45 個⁵：

打敗、播音、裁軍、成交、出席、定點、犯渾、會客、建交、節能、結業、進口、救災、據悉、抗戰、曠課、廉政、留念、漏稅、賣國、命名、命題、簽證、喬裝、請願、缺席、入境、入口、失事、失效、逃荒、挑戰、跳遠、通商、同屋、吸毒、押韻、與會、越冬、造句、整風、值班、酌情、走私、坐班

3.2 離合詞的鑒定及其離析情況考察統計分析

3.2.1 離合詞的甄別鑒定

實際上，這 392 個可離析雙音結構中並非都是離合詞，有一部分結構屬於片語或短語。比如，《大綱》對一些常用結構也不標注詞性。《大綱》指出“爲了便於教學需要，對常用的動賓、動補結構，整體選入，不標注詞性，……”此外，《大綱》對離合詞或常用結構的認定也值得商榷（如《大綱》認爲“跳舞”屬於常用結構）。因而，爲了保證研究的科學性，我們還要對離合詞的範

⁵ 我們又通過國家語委語料庫（約 2000 萬字）進行驗證，結果跟北大語料庫基本相同。未出現離析情況的雙音結構中有少數在國家語委語料庫中出現，它們是：進口（1 例）、失效（2 例）、挑戰（2 例）、吸毒（2 例）、坐班（1 例）、打敗（1 例）。不過，爲了保持統計的一致性，我們仍以北大語料庫的統計資料爲准。

圍進行界定。

3.2.1.1 述補雙音結構

經考察發現，392 個雙音結構的結構大體可分為述補結構和述賓結構（《大綱》中沒有出現主謂等其他結構）。述補結構共有 30 個⁶，如提高、看見、下去、遇見等。這一部分中間一般插入形式固定——只插入“得/不”，實際上是述補結構的可能式或結果式。述補雙音結構的意義一般是前後項 A 義和 B 義的相加，這一部分結構更接近片語。齊滬揚（2000）認為述補式應排除於離合詞之外，鄒立志等（2007）也指出：補充式（即述補式——筆者注）不宜看作離合詞，因為加“得、不”是補充式的標誌性特徵，且不能擴展。

3.2.1.2 述賓凝合雙音詞與述賓組合雙音詞

我們將研究的重點主要放在述賓結構上，這部分更加複雜。這部分可以分為兩類：一類結構緊密，意義抽象，我們一般認為這些是詞，如：洗澡、睡覺、吃虧等；另一類結構鬆散，意義具體，基本上是前後兩項意義的相加，這類結構接近片語或本身就是片語，如：幹活、辦事、送禮、降價等。我們將前者稱之為凝合式雙音結構，後者稱之為組合式雙音結構。由此，我們可以看出，述賓雙音結構家族內部並不統一，而是處於片語與詞之間，構成連續統。

以上是理論上對述賓雙音結構的劃分，具體操作起來還有一定的難度。比如：結構緊密到什麼程度才是凝合式結構，鬆散到什麼程度才是組合式結構？具體到這些雙音結構，哪些是凝合式結構，哪些屬於組合式結構，其鑒定恐怕應該有一個權威、統一而又可操作的標準。

3.2.1.3 凝合結構與組合結構的鑒定標準

《詞典》對條目進行了詞與語素的區分：“在區分詞與非詞的基礎上給單字條目、多字條目標注詞類”。“單字條目在現代漢語中成詞的標注詞類，不成詞的語素和非語素字不做任何標注”（參見《詞典》總目第 6 頁）。我們依據《詞典》確定如下標準⁷：

⁶ 30 個述補結構中有一個述補結構——“打敗”在 CCL 語料庫中沒有發現離析形式，“打敗”在國家語委語料庫僅發現 1 例離析形式。

⁷ 《現代漢語詞典》的認定是否完全準確無誤，我們暫不討論。我們不想陷入語素和詞的界定的爭論中。《現代漢語詞典》是目前最為權威的現代漢語規範詞典，依

- a. 離析雙音結構 A 和 B 二者的義項條目在《詞典》中都被認定為語素的話，那麼該結構就是凝合式結構。如：及格、理髮、註冊、幽默、道歉。
- b. 離析雙音結構 A 和 B 在《詞典》中有一者的義項條目被認定為語素的話，那麼該結構也是凝合式離合結構。例如：洗澡、冒險等。
- c. 離析雙音結構 A 和 B 的義項條目在《詞典》中都被認定為詞，但意義抽象，不是 A 義和 B 義的相加，該結構也是凝合結構。如“拐彎”：例句（3）中“拐彎”的就是組合式；例句（4）中的“拐彎”則為凝合式：
（3）瑞宣呆呆的看著他的後影，直到野求[拐了彎]。【四世同堂】
（4）辣辣腦子[拐了一個彎]，找老朱頭弄了醫院的證明。【你是一條河】
類似的情況還有握手、念書、上臺、下臺等。
- d. 離析雙音結構 A 和 B 的義項條目在《詞典》中都被認定為詞，其意義只是 A 和 B 意義的簡單加合，那麼，該結構就屬組合式。如：訂貨、幹活、辦事、降價等。

3.2.1.4 離合詞

呂叔湘（1979）認為“有些組合只有單一的意義，難於把這個意義分割開來交給這個組合的形式，例如‘走路|洗澡|睡覺|吵架|打仗’等等，因此有人主張管這種組合叫離合詞。”呂先生的見解非常精闢，實際上他所指的“這種組合”就是可以離析的“詞”，也就是我們劃分出來的凝合式述賓離合結構。因此，我們認為，凝合式述賓離析結構符合我們對離合詞的界定⁸，這些結構應該屬於離合詞。

經過分析甄別，392 個雙音結構中，有 45 個雙音結構沒有出現離析現象（44 個述賓結構，1 個述補結構），我們認為，這些結構不具備離合詞的典型特徵，我們姑且不將其作為重點考察物件；還有 29 個述補結構也不宜看作離合詞（見上文分析）。經過分析鑒定，組合式雙音結構 111 個，如：幹活、辦事、送禮、降價等，它們從意義上基本是 A 和 B 的簡單相加，因此接近或屬於片語。這樣我們認定的離合詞就只有凝合式述賓結構 207 個（見全文後附錄一）。請看下表：

據之作爲標準當屬首選。

⁸ 趙元任（1979）：“有一個形式是黏著的，當然整個結構不可能是短語”。

表 3 392 個雙音結構分類表

結構類型 數 值	無離析結構		述補式結構	述賓式離析結構	
	述賓	述補		組合式	凝合式 (離合詞)
數 量	44	1	29	111	207
占總量 (392) 的比率	11.5%		7.4%	28.3%	52.8%

應該特別指出的是，漢語中，詞和片語（短語）構成一個連續統，無法也不可能截然分開，這是漢語的事實。詞和片語/短語界限的模糊性，不僅漢語是這樣，其他語言也是如此。布龍菲爾德指出：“語言中有許多形式是處在黏附形式與詞之間或詞語短語之間的界線上，不可能把哪些形式可以放在絕對的位置上說，或哪些形式不可以，加以硬性的區別。這其中，短語和有些複合詞最易當作邊緣地帶，因為它們的直接形式裏沒有黏附形式”⁹。因此，我們以《大綱》為範圍、以《詞典》為標準認定的離合詞，其與詞和組合式雙音結構的界限也並非涇渭分明，而且我們認定的離合詞內部也有典型與不典型之分。

不過，在科學研究上，以一個相對權威客觀的標準，截取一個相對同質的範圍為研究物件進行研究，是通行的做法。

3.2.2 離合詞的離析情況統計

3.2.2.1 離析頻率分析

上文我們在統計可離析雙音結構時，依據語料中出現的離析形式句數量來反映其離析情況，這種統計可以顯示某結構的離析情況，但還不準確。因為可能有的結構在某些特定情況下出現頻率奇高，那麼一般來說，該結構離析的機會自然就多。所以我們對離合詞離析情況進行統計分析時，為了更加準確地反映離合詞的離析情況，還要分析統計每個離合詞的離析頻率。離析頻率公式如下¹⁰：

$$\text{離合詞的離析頻率 (F)} = \frac{\text{離析例數}}{\text{有效例句總量}}$$

⁹ 見布龍菲爾德，袁家驊譯（1997），《語言論》，212 頁。商務印書館。

¹⁰ 本計算公式參照任海波等（2005）。

離析例數就是我們從語料中得到的含有離合詞離析形式的例句數。所謂有效例句總量，是指出現該離合詞的有效例句量和出現該離合詞離析形式的例句量之和。

所謂離合詞的有效例句是指離合詞處於陳述功能狀態下的句子。離合詞在句法功能上不處於陳述狀態下的例句量，我們就稱之為無效量。比如：（1）“他們在游泳”中的“游泳”在句子中處於陳述狀態，該句就是有效例句；而（2）“他們去了游泳池”中的“游泳”在句法上不處於陳述狀態，這些句子就屬於無效例句。

這樣，我們就可以得出“游泳”的離析頻率：“游泳”出現的離析形式句數是 21 個，有效總量是 599 個。那麼“游泳”的離析頻率為： $F=21/599=0.035$ 。按此方法，我們得出各離合詞的離析頻率。

3.2.2.2 離合詞離析頻率分佈特點

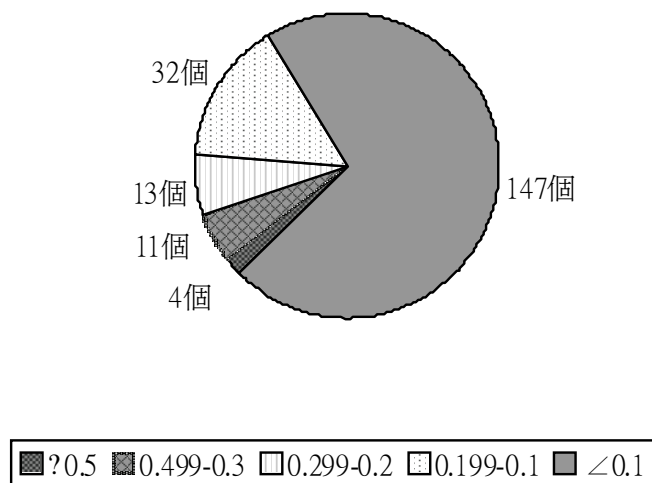
離合詞家族各成員離析頻率不一樣，有些離析頻率高，有些離析頻率低甚至基本不產生離析現象。

我們看 207 個產生離析現象的離合詞，這些離合詞的平均離析頻率僅有 0.059，離析頻率超過 0.5 的只有 4 個。207 個離合詞中 72.9% 的離合詞離析頻率低於 0.1，87.9% 的離合詞其離析頻率低於 0.2。由此可見，在實際語言中絕大部分離合詞（占 98.1%）“合”大於“離”。請看表 4：

表 4 離合詞離析頻率分佈表

離析頻率	≥ 0.5	0.499-0.3	0.299-0.2	0.199-0.1	< 0.1
數量（207）	4	11	13	32	147
比率 （100%）	1.9%	5.3%	6.3%	15.5%	71%

圖1. 207個離合詞離析頻率分配圖



爲了更加直觀，我們製成圖 1。

3.2.2.3 重點離合詞和核心離合詞

用典型範疇理論來說，範疇裏最大量出現的實例總是例示（*instantiate*）了該範疇的原型（*prototype*）。Sinclair J.M &Renouf（1988）也指出“頻率最高的那些詞用途廣泛，並且它們的用法覆蓋了主要的語法點”。應該說，那些高離析頻率和高使用頻率的離合詞代表了離合詞的特點，可以作爲我們教學科研的重要依據（詳見王海峰 2008）。而那些低頻離合詞的代表性較差，不具備典型性。

3.2.2.3.1 重點離合詞

考察發現，只有 60 個離合詞離析頻率超過 0.1，我們稱之爲重點離合詞¹¹。請看下表：

¹¹ 0.1 的離析頻率應該屬於偏低的了。但由於離合詞的離析頻率普遍偏低，我們只能以 0.1 爲最低限。

表 5 60 個重點離合詞離析情況一覽表¹²

序號	離合詞	詞級	離析例數	有效總量	離析頻率
1	歎氣	丙	1, 527	1, 876	0.814
2	搞鬼	丁	31	43	0.721
3	沾光	丁	120	172	0.698
4	聽話	丁	489	911	0.537
5	把關	丁	361	731	0.493
6	放假	甲	185	440	0.42
7	當面	丙	451	1, 131	0.399
8	乾杯	乙	75	203	0.369
9	吃虧	丙	312	987	0.361
10	出名	丁	256	771	0.332
11	鞠躬	丁	225	685	0.328
12	洩氣	丁	68	213	0.319
13	吃苦	丙	309	992	0.311
14	請假	甲	172	563	0.306
15	洗澡	甲	310	1, 024	0.302
16	接班*	丁	36	121	0.298
17	幫忙	乙	642	2, 164	0.297
18	爭氣	丁	108	369	0.293
19	吃驚	乙	528	1, 895	0.279
20	見面	甲	897	4, 164	0.275
21	問好	甲	85	323	0.263
22	打針	乙	61	234	0.261
23	勞駕	甲	28	109	0.257
24	上當	乙	239	966	0.247
25	睡覺	甲	701	2, 847	0.246
26	拐彎*	丙	21	86	0.244
27	操心	丙	128	544	0.235
28	念書*	丙	177	885	0.2
29	隨便	乙	80	407	0.197
30	照相	甲	97	509	0.191

¹² 表中離合詞帶有“*”符號的，請見本表後意義解釋。以後各章節引用這些離合詞用例，除非特別說明，均指該表所釋的意義。

31	排隊	丁	185	975	0.19
32	告狀	丁	69	369	0.187
33	敬禮	乙	72	387	0.186
34	出神	丁	63	338	0.186
35	吵架	丙	117	631	0.185
36	拜年	丁	43	239	0.18
37	狠心	丁	55	321	0.171
38	遭殃	丁	27	150	0.16
39	開課	乙	14	85	0.165
40	打架	丙	142	932	0.152
41	鼓掌	乙	121	808	0.15
42	交手	丁	29	195	0.149
43	下臺 [*]	丁	64	448	0.143
44	吵嘴	丁	24	172	0.14
45	留神	丁	54	393	0.137
46	翻身 [*]	丙	29	211	0.137
47	插嘴	丁	34	255	0.133
48	化妝	丁	80	600	0.133
49	丟人	丁	39	299	0.13
50	走道 [*]	乙	8	62	0.13
51	結婚	乙	678	5, 530	0.123
52	跳舞	甲	139	1, 137	0.122
53	做主	丁	28	229	0.122
54	站崗	丁	33	274	0.12
55	隨意	丁	9	75	0.12
56	生氣	乙	262	2, 267	0.115
57	理髮	乙	45	399	0.112
58	掛鈎 [*]	丁	90	837	0.108
59	倒楣	丙	62	591	0.105
60	傷心	乙	164	1, 635	0.10

說明（依據《現代漢語詞典》第 5 版）：

“接班”指繼承前人的工作或事業。“拐彎”指（思路、語言等）轉變方向。“念書”指讀書或學習。“下臺”比喻擺脫困難窘迫的處境。“翻身”指改變落後面貌或不利局面。“走道”指走路、行走。“掛鈎”指建立某種聯繫。

3.2.2.3.2 核心離合詞

上表中 60 個離合詞離析頻率雖然超過 0.1，但有的離析例數和有效總量偏低。如“搞鬼”，雖然其離析頻率高達 0.721，但語料庫中出現的離析例數和有效總量僅為 31/43。“走道”、“隨意”也是如此，雖然離析頻率超過 0.1，但語料庫中出現的離析例數和有效總量僅為 8/63 和 9/75，說明這些詞使用頻率較低。

我們確定離析頻率超過 0.1，且離析例數在 200 句以上的 16 個離合詞為核心離合詞，應該說這 16 個核心離合詞的典型性更強。這 16 個核心離合詞是：

歎氣、聽話、把關、當面、吃虧、出名、鞠躬、吃苦、洗澡、幫忙、吃驚、見面、上當、睡覺、結婚、生氣。

3.2.3 離合詞離析形式考察

我們通過CCL語料庫對圈定的207個《大綱》中的離合詞離析形式進行了窮盡的考察，得到含離合詞離析形式的離析形式句14,968句。通過對207個離合詞的離析形式進行細緻的歸納整理，我們發現離合詞在實際語料中的離析形式可以大致歸納為13種，以下表7和表8分別是從不同角度統計出的結果。

表7 207個離合詞離析形式一覽表¹³

位次	離析形式	出現該離析形式的離合詞的數量	出現該離析形式離合詞所占（207個）的比率	舉例 （以“幫忙”等為例）
1	A+了（+其他形式）+B ¹⁴	160	77.29%	幫了忙、幫了一個忙、幫了他的忙
2	A+補語 ¹⁵ +B	103	49.76%	幫完忙、幫起忙來
3	A+名詞／代詞（的）+B	84	40.58%	幫老師忙、幫他的忙

¹³ 每種離析形式的詳細情況請見文後附錄二。

¹⁴ 符號“A”代表離合詞的前項形式，“B”代表後項形式，以下同。

¹⁵ 數量詞分名量詞、動量詞、時量詞等，它們在語法結構中功能不同：有的修飾名詞，有的修飾補充動詞。但是，它們作為離合詞 AB 的離析形式，具有一致性，就是它們都不是單獨同 A 或 B 發生關係，而是跟 AB 整個結構發生關係。為了保持分類的整齊，我們在本表中暫將其歸為一類，表 8 與此一致。它們各自的詳細分類和資料請見王海峰（2008）。

4	A+數量詞 ¹⁶ +B	84	40.58%	幫一幫忙、幫兩個忙、幫一點忙
5	A+過（+其他形式）+B	74	35.75%	幫過忙、幫過他的忙、幫過一次忙、幫過大忙
6	前置 B+A	57	27.5%	忙幫了、連忙都不幫、把字簽了
7	A+著（+其他形式）+B	41	19.81%	幫著忙、幫著他的忙
8	A+的+B	36	17.39%	幫的忙
9	A+個 ¹⁷ +B	34	16.43%	幫個忙
10	A+形容詞 ¹⁸ （+的）+B	32	15.46%	幫大忙
11	A重疊+B	19	9.18%	幫幫忙、鼓了鼓掌、伸一伸手
12	A+數詞+B	13	6.27%	睡一覺 見兩面
13	A+動詞性形式（+的）+B	9	4.35%	吃管閒事的虧

說明：

甲.某一離合詞可以出現多種離析形式，所以出現該離析形式離合詞所占（207個）的比率總和超過100%。表中的離析形式並不是嚴格的分類標準，只是對離析形式特徵的概括。表8與此一致。

乙.A+了（+其他形式）+B，包括：A+了+B、A+了+數量詞+B、A+了+數量

¹⁶ 數量詞分名量詞、動量詞、時量詞等，它們在語法結構中功能不同：有的修飾名詞，有的修飾補充動詞。但是，它們作為離合詞 AB 的離析形式，具有一致性，就是它們都不是單獨同 A 或 B 發生關係，而是跟 AB 整個結構發生關係。為了保持分類的整齊，我們在本表中暫將其歸為一類，表 8 與此一致。它們各自的詳細分類和資料請見王海峰（2008）。

¹⁷ 離合詞中間插入“個”，如“幫個忙”，其中“個”跟“吃了兩個饅頭”的“個”功能不同（見後文分析）。所以，我們不將“個”歸入量詞，而將“A+個+B”形式單列出來。表 8 與此一致。

¹⁸ 指表層語法形式上修飾 B、做定語的那部分形容詞。本文其他表格及文字表述部分與此一致。

- 詞+名詞+B、A+了+數量詞+形容詞+B、A+了+數詞+B、A+了+名詞/代詞(的)+B、A+了+名詞/代詞+形容詞+B等。
- 丙.A+過(+其他形式)+B，包括：A+過+B、A+過+數量詞+B、A+過+數詞+B、A+過+名詞/代詞+B、A+過+形容詞(的)+B、A+過+動詞+B等。
- 丁.A+著(+其他形式)+B，包括：A+著+B、A+著+名詞/代詞(的)+B、A+著+量詞+B、A+著+形容詞+B等。
- 戊.在統計離析形式時，爲了避免重複計算，我們僅計算第一出現插入形式。如：A+了+數量詞+B，其離析形式數量只計入A+了(+其他形式)+B中，不再計入A+數量詞+B中。表8的計算方法跟此一致。

爲了更加準確起見，我們還對60個重點離合詞中出現的11,498條離析形式句按出現某離析形式的例數進行統計，請看下表：

表8 60個重點離合詞離析形式（按出現某離析形式的例數統計）

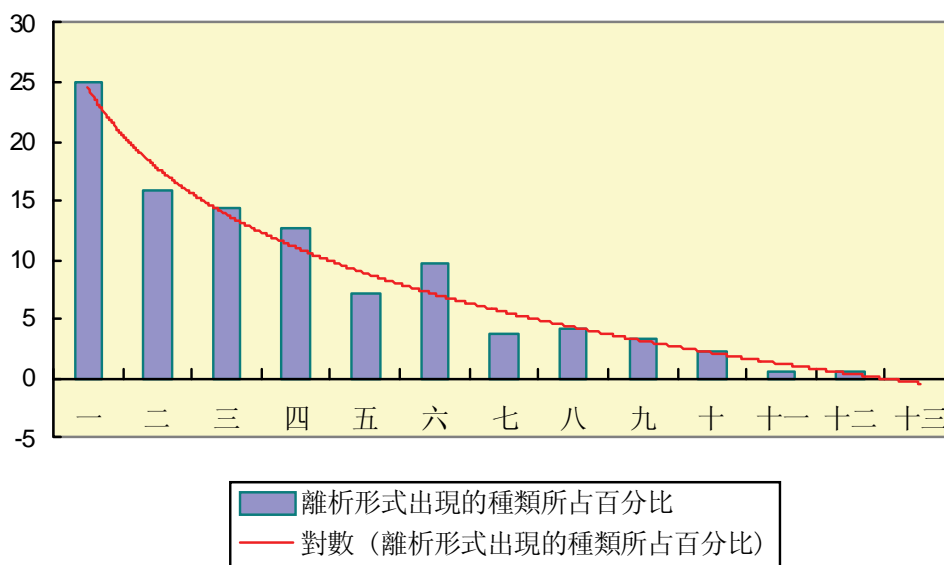
位次	離析形式	該形式出現例數（共11,498句）	所占百分比
1	A+了（+其他形式）+B	4,253	37%
2	A+名詞／代詞(的)+B	1,518	13.20%
3	A+補語+B	1,274	11.80%
4	A+數量詞+B	1,033	8.98%
5	A+過（+其他形式）+B	869	7.56%
6	A+著（+其他形式）+B	592	5.15%
7	前置 B+A	316	2.75%
8	A+數詞+B	292	2.54%
9	A+形容詞（+的）+B	275	2.39%
10	A+個+B	293	2.55%
11	A重疊+B	161	1.4%
12	A+的+B	111	0.97%
13	A+動詞性形式（+的）+B	13	0.11%

綜合表7和表8，我們發現，這兩個表離合詞離析形式出現的位次出入不大。兩表都顯示離合詞插入“了”的離析形式最多，其次是插入補語或插入名詞/代詞以及數量詞的離析形式，然後是插入體標記“過”，再次是插入“著”和B前置，這七種離析形式應該是離合詞的主要離析形式。離合詞中插入數詞、形容詞、“個”、“的”以及動詞A重疊的情況較居後位元，而插入動詞性形式的情況最少，居末位元。

3.2.4 離合詞離析形式出現種類情況考察

我們發現，雖然 207 個離合詞出現的離析形式有 13 種，但是，這些離合詞離析形式出現的種類分佈並不一致。統計發現，離合詞離析形式的出現種類與以全部這些形式離析的離合詞的數量和所占比例成反比：越是離析形式種類少的離合詞，其數量和所占比例就越大；反之亦然。請看下圖：

圖 2 離合詞離析形式情況圖



統計還發現，大多數離合詞出現的離析形式種類在四種以下，占到 68.1%，出現五種以上（含五種）離析形式的離合詞，只占 31.9%；出現七種以上（含七種）離析形式的離合詞，只占 15%；出現九種以上（含九種）離析形式的離合詞，只占 6.8%；而出現十一種以上（含十一種）離析形式的離合詞，則不到 1%。

由此，我們結合上文 3.2.2 統計結果認為，無論從離合詞離析頻率分佈特點來看，還是從離析形式分佈特點來看，都可以得出以下結論：所謂離合詞，實際上“合”是“常態”（regular），“離”是“異態”（idiosyncratic）。

3.3 離合詞離析頻率與使用頻率的關係考察

我們發現，離合詞離析頻率與其在語言中的使用頻率有一定關係。大部分高離析頻率離合詞常常也是現實生活中高使用頻率的詞語。低使用頻率離合詞也常常是低離析頻率離合詞。請看下表（其中使用頻率資料依據的是《現代漢語頻率詞典》（北京語言學院出版社 1986））：

表 9 前 15 位高離析率離合詞與後 15 位低離析率離合詞使用頻率對比表

離合詞	離析頻率	使用頻率	離合詞	離析頻率	使用頻率
睡覺	0.246	0.01164	節能	0 ¹⁹	0.00107
生氣	0.115	0.00768	賣國	0	0.00084
見面	0.275	0.00662	命名	0	0.00084
幫忙	0.297	0.00555	移民	0	0.00061
結婚	0.123	0.00517	挑戰	0	0.00084
吃驚	0.279	0.00502	成交	0	0.00076
吃虧	0.361	0.00274	請願	0	0.00068
歎氣	0.814	0.00228	建交	0	0.00068
上當	0.247	0.00183	逃荒	0	0.00061
洗澡	0.302	0.00175	會客	0	0.00038
當面	0.399	0.00167	走私	0	0.00038
鞠躬	0.328	0.00167	與會	0	0.0003
吃苦	0.311	0.00114	失效	0	0.0003
聽話	0.537	0.00099	結業	0	0.0003
出名	0.332	0.00038	命題	0	0.0003
平均	0.333	0.003742	平均	0	0.000593

上表顯示離析頻率與使用頻率基本正相關，高使用頻率離合詞的離析頻率高，低使用頻率離合詞離析頻率低。對比來看，高離析頻率離合詞的平均使用頻率是零離析率離合詞平均使用頻率的 6 倍多。

19 CCL 語料庫中零離析頻率離合詞並不說明該離合詞在語言中沒有離析的可能，只說明該離合詞的離析頻率很低。如：在我們語料中未出現離析情況的雙音結構中有少數在國家語委語料庫中低頻出現，它們是：進口（1 例）、失效（2 例）、挑戰（2 例）、吸毒（2 例）、坐班（1 例）。因此零離析頻率離合詞不等於非離合詞。

4. 結語

以上是根據大型語料庫對離合詞考察後得出的部分結果，限於篇幅，其他考察結果和資料恕不一一列出，針對這些考察結果所進行的深入闡述也只好留待後文盡述。

就筆者所知，由語料庫入手從對外漢語教學角度研究離合詞還是比較鮮見的。值得一提的是，任海波、王剛（2005）此前利用語料庫對離合詞形式作了一些初步的探索。

但是任、王的探索不無缺憾：首先，他們的語料庫很小，只有 1300 多萬字。其次，他們考察的語料僅限於小說。在語料量和語體分佈極其有限的情況下，能否全面探求離合詞的實質，值得商榷。正如任、王坦言：“反映的只是這些離合詞在受限語料（即當代文學小說—作者注）的表現，體現出來的離合頻度的不同也只是一種相對的差異”。導致缺少離散形式表現的原因“一是語料規模的限制”；“二是語體範圍限制”。

此外，他們沒有區分離析雙音結構的不同形式，而將一些本來不能算作離合詞的結構納入進來，這樣就影響了結論的嚴密性與科學性。如：他們得出的典型離合詞是：聽話、分清、出事、吸煙、上當、乾杯、吃驚、打仗、吃虧、犯罪、握手、享福。且不說這些離合詞的典型程度是否跟說漢語的人的語感相符，更重要的是，有的本不是詞的結構被當成了典型離合詞，如：吸煙，連收詞略寬的《現代漢語詞典》（2005 年第 5 版）都沒有收之為詞，任、王將其定為典型離合詞，難以令人信服。

實際上，任、王自己也充滿疑惑：“這 12 個典型離合詞中有幾個可能在定性時引起爭議。以‘吸煙’為例，‘吸’和‘煙’都是能夠單說或至少是可以獨立成詞的形式（語素），因此如果按照美國結構主義對合成短語的定義，那麼這裏的‘吸煙’就可能被確定為合成短語，而不是（離合）詞。此外‘分清’、‘握手’等也多多少少存在類似的問題。”如果認定的典型有三分之一以上有問題，那麼典型就會大打折扣。可能就不只是像作者所說的“這種情況的存在儘管可能會影響到典型離合詞數量的多寡，但不影響其他離合詞作為離合詞典型成員的資格”。

我們通過大型語料庫嘗試從科學的角度出發，對離合詞的離析形式等各方面的情況進行了更全面的考查，發現了許多有意思的現象。

我們研究的特點是：

（1）利用大型平衡語料庫對離合詞進行系統研究，保證結論的客觀性；

(2) 嚴格區分不同性質的離析結構，以確保研究的科學性；(3) 對離合詞進行考察，糾正了人們長期以來關於離合詞的部分不正確的認識；(4) 本研究有明確的目的性，強調實用，為對外漢語教學和研究提供資料支援。

附錄

附錄一 207 個離合詞

見面	幫忙	吃驚	聽話	當面	把關	吃苦	吃虧	生氣
出名	上當	鞠躬	傷心	打架	操心	跳舞	鼓掌	吵架
沾光	爭氣	掛鈎	問好	念書	化妝	隨便	下臺	乾杯
敬禮	告狀	洩氣	出神	倒楣	打針	狠心	理髮	留神
丟人	掛號	接班	插嘴	搞鬼	站崗	交手	注意	翻身
勞駕	作主	遭殃	拐彎	吵嘴	沒轍	隨意	走道兒	歎氣
澡洗	睡覺	結婚	押韻	開口	分期	失約	冒險	懷孕
帶頭	講理	吹牛	開刀	上學	談天	防汛	離婚	參軍
盡力	聊天	放心	要命	報名	說情	請客	抽空	退休
著急	放學	灰心	拼命	出院	起床	讀書	讓步	道歉
動手	搗亂	剪綵	游泳	出差	用心	作客	招生	造反
起哄	提醒	住院	稱心	加油	曠工	辭職	種地	用功
通風	算數	失業	畢業	跑步	散步	動身	放手	埋頭
作案	有名	打獵	懂事	防疫	升學	納悶	伸手	宣誓
像樣	及格	待業	致電	發炎	滑冰	罷工	入學	搗蛋
投標	對頭	上臺	閉幕	減產	作文	就職	擔心	滿月
破產	報銷	失蹤	登陸	致富	完蛋	失學	開學	考試
探親	決口	害羞	登記	保密	貶值	開工	分紅	滑雪
投資	幽默	延期	生效	定性	報到	辦公	作戰	集郵
還原	留意	安心	出面	絕望	行軍	提名	迎面	行賄
偷稅	值勤	執政	起身	開幕	註冊	成套	到期	違法
辦學	配套	送行	起草	獻身	加工	上任	投產	結果
集資	移民	增產	出門	就業	攝影	施工	出口	招手

附錄二離合詞插入形式情況²⁰

1. 插入“了/了+其他形式”的離合詞

這些離合詞最多：

見面 幫忙 吃驚 聽話 當面 吃苦 吃虧 生氣 出名 上當 鞠躬 傷心 打架
操心 跳舞 鼓掌 吵架 沾光 爭氣 掛鉤 問好 念書 化妝 下臺 乾杯 敬禮
告狀 洩氣 出神 倒楣 打針 狠心 理髮 留神 丟人 掛號 接班 插嘴 搞鬼
站崗 交手 翻身 作主 遭殃 拐彎 吵嘴 排隊 走道 歎氣 洗澡 睡覺 結婚
開口 失約 冒險 懷孕 帶頭 吹牛 開刀 上學 談天 離婚 參軍 照相 握手
請假 拜年 聊天 放心 要命 報名 說情 請客 抽空 退休 著急 放學 灰心
拼命 出院 起床 讀書 讓步 道歉 動手 搗亂 剪綵 游泳 出差 作客 造反
起哄 提醒 住院 稱心 加油 辭職 種地 失業 畢業 散步 動手 作案 有名
打獵 懂事 納悶 伸手 宣誓 致電 發炎 罷工 入學 投標 對頭 上臺 閉幕
減產 作文 就職 擔心 滿月 破產 失蹤 登陸 致富 完蛋 失學 開學 考試
決口 害羞 登記 貶值 開工 投資 幽默 延期 生效 定性 報到 還原 出面
絕望 提名 迎面 行賄 起身 開幕 註冊 到期 辦學 起草 加工 上任 結果
移民 出門 放假 招手

1) 插入“了”或“了+其他形式”形式的共有 160 個，占總數的 77.29%

①僅插入“了”的離合詞有 78 個。

②既插入了“了”又插入“了+其他形式”的離合詞有 63 個。

③僅插入了“了+其他形式”的有 19 個。

2) 其中“了”後的形式多於三種形式的有 25 個：

見面 幫忙 吃苦 吃虧 生氣 上當 鞠躬 操心 跳舞 吵架 沾光 念書 乾杯
敬禮 告狀 倒楣 拐彎 吵嘴 歎氣 洗澡 睡覺 結婚 帶頭 請客 讀書

這些離合詞離析情況比較複雜，學生掌握起來有難度。

2. 插入“著 / 著+其他形式”的離合詞

這些離合詞有：

當面 生氣 鞠躬 傷心 打架 操心 跳舞 鼓掌 沾光 問好 念書 化妝 隨便
出神 狠心 留神 拐彎 歎氣 洗澡 睡覺 冒險 懷孕 上學 談天 聊天 放心
拼命 讀書 道歉 做客 起哄 散步 埋頭 納悶 擔心 保密 排隊 握手 請假
迎面 招手

²⁰ 本部分由薛晶晶博士協助整理統計，謹致之謝忱。

插入“著”形式的離合詞，共 41 個，占總數的 19.81%。

①僅插入“著”的離合詞，共 29 個。

②僅插入“著+其他形式”的離合詞，共 1 個。

③既插入了“著”又插入“著+其他形式”的離合詞，共 12 個。

3. 插入“過/過+其他形式”的離合詞

這些離合詞有：

見面 幫忙 吃苦 吃虧 生氣 上當 鞠躬 傷心 打架 操心 跳舞 吵架 沾光
問好 念書 化妝 倒楣 打針 理髮 站崗 交手 吵嘴 洗澡 睡覺 結婚 押韻
開口 冒險 懷孕 開刀 上學 離婚 參軍 照相 握手 請假 拜年 聊天 請客
著急 起床 讀書 讓步 道歉 動手 搗亂 游泳 用心 做客 造反 住院 種地
用功 算數 失業 畢業 散步 作案 打獵 納悶 伸手 宣誓 及格 待業 入學
失學 探親 登記 作戰 集郵 註冊 加工 出門 招手

插入形式“過”或“過+其他形式”的離合詞有 74 個，占總數（207）的 35.75%。

①僅插入“過”的離合詞有 34 個。

②僅插入“過+其他形式”的離合詞有 7 個。

③既插入“過”又插入“過+其他形式”的離合詞有 33 個。

4. 插入補語的離合詞

這些離合詞有：

見面 幫忙 把關 吃苦 吃虧 生氣 鞠躬 傷心 打架 操心 跳舞 鼓掌 吵架
沾光 爭氣 掛鈎 問好 念書 化妝 乾杯 敬禮 出神 倒楣 打針 狠心 理髮
丟人 掛號 接班 插嘴 站崗 交手 翻身 勞駕 作主 拐彎 吵嘴 走道兒 歎氣
洗澡 睡覺 結婚 開口 懷孕 帶頭 講理 吹牛 開刀 上學 談天 離婚 照相
握手 請假 放假 拜年 聊天 放心 請客 抽空 著急 拼命 出院 起床 讀書
動手 游泳 住院 加油 辭職 用功 算數 畢業 跑步 散步 動身 放手 埋頭
作案 開學 伸手 宣誓 罷工 入學 減產 作文 擔心 報銷 致富 考試 探親
登記 保密 開工 投資 幽默 辦公 投產 結果 出門 施工 排隊

插入補語形式離合詞有 103 個，占總數（207）的 49.76%。

5. 插入“個”的離合詞

這些離合詞有：

見面 幫忙 出名 鞠躬 生氣 打架 跳舞 掛鈎 敬禮 理髮 掛號 翻身 作主
洗澡 拐彎 睡覺 結婚 冒險 帶頭 說情 請客 抽空 道歉 提醒 算數 散步
有名 登記 出面 照相 請假 排隊 放假 拜年

插入“個”的離合詞有 34 個，占總數（207）的 16.43%。

6. 插入數詞的離合詞

1) “一”

見面 吃驚 鞠躬 打架 吵架 乾杯 打針 睡覺 開口 開刀 攝影

插入“一”的離合詞共 11 個，占總數（207）的 5.31%。

2) 其他數詞“兩”、“幾”等

見面 上當 鞠躬

插入其他數字的離合詞共 3 個，占總數的 1.45%。

7. 插入數量詞（動量、名量、時量、不定量等）的離合詞

這些離合詞有：

見面 幫忙 吃苦 吃虧 生氣 上當 鞠躬 打架 操心 跳舞 鼓掌 吵架 沾光
爭氣 問好 念書 敬禮 倒楣 打針 狠心 理髮 留神 丟人 掛號 插嘴 站崗
注意 作主 拐彎 吵嘴 走道兒 歎氣 洗澡 結婚 押韻 開口 冒險 懷孕 帶
頭 講理 吹牛 上學 離婚 照相 握手 請假 排隊 放假 拜年 聊天 放心 報
名 請客 著急 放學 灰心 起床 讀書 讓步 道歉 動手 搗亂 游泳 出差 起
哄 住院 加油 曠工 用功 跑步 散步 懂事 防疫 伸手 搗蛋 作文 擔心 投
資 行軍 加工 集資 增產 出門 招手

共 84 個，占總數（207）的 40.58%。

8. 插入名詞（的）/ 代詞（的）的離合詞

這些離合詞有：

見面 幫忙 聽話 當面 把關 吃苦 吃虧 生氣 上當 傷心 打架 操心 跳舞
鼓掌 吵架 沾光 爭氣 問好 念書 化妝 隨便 乾杯 告狀 洩氣 出神 倒楣
丟人 掛號 接班 插嘴 搞鬼 站崗 勞駕 作主 吵嘴 照相 排隊 隨意 走道
兒 歎氣 洗澡 睡覺 結婚 冒險 講理 吹牛 開刀 上學 離婚 聊天 要命 請
客 著急 放學 拼命 讀書 道歉 搗亂 游泳 出差 用心 造反 起哄 提醒 稱

心 辭職 用功 算數 打獵 懂事 致電 發炎 罷工 作文 擔心 破產 害羞 保密 投資 辦學 送行 獻身 結果 就業
共 84 個，占總數（207）的 40.58%。

9. 插入形容詞（的）的離合詞

這些離合詞有：

吃驚 吃苦 吃虧 生氣 上當 鞠躬 打架 操心 鼓掌 爭氣 念書 化妝 敬禮 倒楣 理髮 站崗 歎氣 睡覺 冒險 談天 防汛 請假 排隊 拜年 聊天 拼命 讀書 出差 開工 生效 作戰 執勤
共 32 個，占總數的 15.46%。

10. 插入動詞結構（的）的離合詞

這些離合詞共 9 個，占總數的 4.35%。

吃苦 吃虧 打架 操心 跳舞 爭氣 冒險 道歉 致電

11. 插入（的）的離合詞

這些離合詞有：

幫忙 吃虧 生氣 鞠躬 打架 操心 跳舞 念書 告狀 理髮 接班 搞鬼 作主 結婚 開口 冒險 帶頭 上學 離婚 參軍 說情 請客 退休 起床 讀書 動手 請假 排隊 放假 動身 作案 上臺 投資 定性 提名 出門
共 36 個，占總數的 17.39%。

12. A 重疊的離合詞

1) AAB

幫忙 出名 傷心 沾光 打針 接班 洗澡 說情 動手 出面 招手 送行 照相
共 13 個，占總數的 6.28%。

2) A 了 AB

鼓掌 化妝 狠心 歎氣 握手
共 5 個，占總數的 2.42%。

3) A-AB

伸手
共 1 個，占總數的 0.48%。

13. B 提前的離合詞

這些離合詞有：

見面 幫忙 把關 吃苦 吃虧 生氣 上當 傷心 操心 跳舞 沾光 爭氣 問好
化妝 告狀 洩氣 理髮 丟人 掛號 插嘴 翻身 洗澡 睡覺 結婚 開口 冒險
懷孕 講理 吹牛 上學 離婚 放心 請客 著急 放學 灰心 拼命 讀書 讓步
游泳 出差 用心 辭職 用功 有名 懂事 排隊 閉幕 考試 保密 開工 辦公
開幕 辦學 招手 請假 排隊

共 57 個，占總數的 27.5%。

附錄三 CCL 語料庫具體數據²¹

（共有 157 個檔，24 個檔夾，229,700,435 位元組，約 1.15 億字）

Name	Type	Size(Byte)
(一)xiandai\當代	path	217800159
xiandai\當代\人民日報（1995、1996、2004）	path	102436767
xiandai\當代\市場報 1994	path	16181356
xiandai\當代\作家文摘（93—97）	path	28201590
xiandai\當代\讀者（合訂本）	path	24459989
xiandai\當代\應用文	path	10143622
xiandai\當代\口語	path	259800
xiandai\當代\文學	path	24888601
1. xiandai\當代\文學\臺灣作家	path	3362894
xiandai\當代\文學\臺灣作家\東尼！東尼！.txt	file	205690
xiandai\當代\文學\臺灣作家\小李飛刀.txt	file	906553
xiandai\當代\文學\臺灣作家\巴西狂歡節.txt	file	225138
xiandai\當代\文學\臺灣作家\月朦朧烏朦朧.txt	file	242842
xiandai\當代\文學\臺灣作家\夢的衣裳.txt	file	201324
xiandai\當代\文學\臺灣作家\玉卿嫂.txt	file	54960
xiandai\當代\文學\臺灣作家\紅頂商人胡雪岩.txt	file	1060507
xiandai\當代\文學\臺灣作家\金大班的最後一夜.txt	file	18006

²¹ 北大漢語言研究中心（CCL）詹衛東博士提供，謹致謝忱。

xiandai\當代\文學\臺灣作家\雁兒在林梢.txt	file	245840
xiandai\當代\文學\臺灣作家\青青河邊草.txt	file	202034
2. xiandai\當代\文學\大陸作家	path	17859322
xiandai\當代\文學\大陸作家\一地雞毛.txt	file	66132
xiandai\當代\文學\大陸作家\一百個人的十年.TXT	file	431341
xiandai\當代\文學\大陸作家\上海寶貝.txt	file	276181
xiandai\當代\文學\大陸作家\余華.TXT	file	1228991
xiandai\當代\文學\大陸作家\余秋雨.TXT	file	369324
xiandai\當代\文學\大陸作家\佳作 1.TXT	file	566249
xiandai\當代\文學\大陸作家\佳作 2.TXT	file	559548
xiandai\當代\文學\大陸作家\佳作 4.TXT	file	547638
xiandai\當代\文學\大陸作家\劉心武短篇.TXT	file	295412
xiandai\當代\文學\大陸作家\劉心武選集.TXT	file	668982
xiandai\當代\文學\大陸作家\北京人在紐約.TXT	file	160426
xiandai\當代\文學\大陸作家\單位. Txt	file	79392
xiandai\當代\文學\大陸作家\周恩來的最後十年.txt	file	168580
xiandai\當代\文學\大陸作家\幹校六記.TXT	file	57730
xiandai\當代\文學\大陸作家\床上的月亮.txt	file	54152
xiandai\當代\文學\大陸作家\張一弓.TXT	file	114849
xiandai\當代\文學\大陸作家\張承志.txt	file	226516
xiandai\當代\文學\大陸作家\張煒.TXT	file	704069
xiandai\當代\文學\大陸作家\張賢亮.TXT	file	465242
xiandai\當代\文學\大陸作家\當代短篇小說 1.TXT	file	782779
xiandai\當代\文學\大陸作家\徐坤.txt	file	64811
xiandai\當代\文學\大陸作家\故鄉天下黃花.txt	file	365775
xiandai\當代\文學\大陸作家\方方.txt	file	365120
xiandai\當代\文學\大陸作家\曾卓文集.txt	file	600540
xiandai\當代\文學\大陸作家\朱自清與中國.txt	file	36711
xiandai\當代\文學\大陸作家\梁曉聲.TXT	file	688579
xiandai\當代\文學\大陸作家\比如女人.txt	file	457778

xiandai\當代\文學\大陸作家\畢淑敏.TXT	file	150631
xiandai\當代\文學\大陸作家\池莉.txt	file	295477
xiandai\當代\文學\大陸作家\汪曾祺.TXT	file	677679
xiandai\當代\文學\大陸作家\王曉波.txt	file	752546
xiandai\當代\文學\大陸作家\王朔 a.txt	file	575193
xiandai\當代\文學\大陸作家\王朔 b.txt	file	604070
xiandai\當代\文學\大陸作家\王朔 c.txt	file	667089
xiandai\當代\文學\大陸作家\王朔 d.txt	file	717233
xiandai\當代\文學\大陸作家\王朔 e.txt	file	374560
xiandai\當代\文學\大陸作家\王蒙.TXT	file	84539
xiandai\當代\文學\大陸作家\甜蜜蜜.txt	file	13167
xiandai\當代\文學\大陸作家\皇城根.TXT	file	671927
xiandai\當代\文學\大陸作家\編輯部的故事.TXT	file	276289
xiandai\當代\文學\大陸作家\老舍長篇 3.txt	file	342442
xiandai\當代\文學\大陸作家\蔣子龍.TXT	file	117810
xiandai\當代\文學\大陸作家\鄧友梅選集.TXT	file	585753
xiandai\當代\文學\大陸作家\鄧賢.TXT	file	183568
xiandai\當代\文學\大陸作家\鐵凝.TXT	file	20203
xiandai\當代\文學\大陸作家\雪白血紅.TXT	file	346299
3. xiandai\當代\文學\香港作家	path	3666385
xiandai\當代\文學\香港作家\闔家歡.txt	file	536801
xiandai\當代\文學\香港作家\激情三百日.txt	file	193112
xiandai\當代\文學\香港作家\白馬嘯西風.TXT	file	138091
xiandai\當代\文學\香港作家\神雕俠侶.txt	file	1855012
xiandai\當代\文學\香港作家\蜜糖兒.txt	file	264549
xiandai\當代\文學\香港作家\豪門驚夢.txt	file	220443
xiandai\當代\文學\香港作家\金融大風暴.txt	file	112486
xiandai\當代\文學\香港作家\雪山飛狐.TXT	file	272736
xiandai\當代\文學\香港作家\青春偶像.txt	file	73155
xiandai\當代\翻譯作品	path	3494073
xiandai\當代\翻譯作品\人腦之迷.TXT	file	151806

xiandai\當代\翻譯作品\大腦如何思維.TXT	file	196495
xiandai\當代\翻譯作品\時間簡史.txt	file	218898
xiandai\當代\翻譯作品\普通心理學.TXT	file	1119932
xiandai\當代\翻譯作品\猜想與反駁.TXT	file	641033
xiandai\當代\翻譯作品\自私的基因.txt	file	341215
xiandai\當代\翻譯作品\資本論.TXT	file	824694
xiandai\當代\詞典	path	7734361
xiandai\當代\詞典\倒序現漢詞典.txt	file	4021932
xiandai\當代\詞典\現漢詞典.txt	file	3712429
(二) xiandai\現代	path	11900157
xiandai\現代\戲劇	path	1062882
xiandai\現代\戲劇\日出.txt	file	92193
xiandai\現代\戲劇\老舍戲劇 1.TXT	file	794636
xiandai\現代\戲劇\雷雨.txt	file	176053
xiandai\現代\文學	path	10837275
xiandai\現代\文學\俞平伯.TXT	file	344132
xiandai\現代\文學\四世同堂.TXT	file	1436208
xiandai\現代\文學\散文 1.TXT	file	1385037
xiandai\現代\文學\散文 2.TXT	file	955345
xiandai\現代\文學\散文 3.TXT	file	1329780
xiandai\現代\文學\柔石.txt	file	26703
xiandai\現代\文學\沈從文.TXT	file	463938
xiandai\現代\文學\現代短篇.TXT	file	370938
xiandai\現代\文學\老舍短篇.TXT	file	972750
xiandai\現代\文學\老舍長篇 1.TXT	file	945708
xiandai\現代\文學\老舍長篇 2.txt	file	1369370
xiandai\現代\文學\茅盾.TXT	file	103118
xiandai\現代\文學\錢鐘書.TXT	file	667618
xiandai\現代\文學\魯迅.TXT	file	466630

參考文獻

- Halliday, M A K & E. (2004) McDonald. Metafunctional Profile of the Grammar of Chinese Caffarel. A. J. R. Martin & C. M. I. M. Matthiessen. Language Typology: A Functional Perspective Amsterdam. Benjamins.
- Hopper, Paul. (1987) Emergent Grammar. Berkeley Linguistic society.
- Hunston & Francis. (2000) Pattern Grammar: A Corpus Driven Approach to the Lexical Grammar of English. John Benjamins.
- Miller D. G. (1993) Complex verb formation, Amsterdam: John benjamins.
- Sinclair, . (1991) Corpus, concordance, collocation: oxford, Oxford university press.
- Sinclair J. M & Renouf A. . (1988) A Lexical Syllabus for Language Learning. In R. Carter & M. McCarthy (eds) Vocabulary and Language Teaching, Harlow, Langman.
- Tognini-Bonelli, E. (2001) Corpus Linguistics at Work. Amsterdam: John Benjamins.
- 丁聲樹等 (1961), 《現代漢語語法講話》。北京: 商務印書館。
- 王 力 (1982), 《漢語語法綱要》。上海: 上海教育出版社。
- 王 寧 (1997), 〈現代漢語雙音合成詞的構詞理據與古今漢語的溝通〉, 《慶祝中國社會科學院語言研究所建所 45 周年學術論文集》, 中國語文編輯部編。北京: 商務印書館。
- 王 勇 (2008), 〈行走在語法和辭彙之間——型式語法述評〉, 《當代語言學》, 第 3 期, 257-266。
- 王洪君 (1994), 〈從字和字組看詞和短語〉, 《中國語文》, 第 2 期, 102-112。
- 王鐵利 (2001), 〈現代漢語離合詞問題研究〉, 中國社會科學院研究生院碩士學位論文。
- 王海峰 (2002), 〈現代漢語離合詞離析動因芻議〉, 《語文研究》, 第 3 期, 29-34。
- 王海峰 (2007), 〈語料庫在對外漢語教學中的作用〉, 《漢語教學學刊》, 第 3 期, 212-227。
- 王海峰 (2008), 《現代漢語離合詞離析形式功能研究》, 北京語言大學博士論文。

- 中國社會科學院語言研究所詞典編輯室（2005），《現代漢語詞典》，第5版。
北京：商務印書館。
- 布龍菲爾德（1997），《語言論》，袁家驊譯。北京：商務印書館。
- 盧偉（1999），〈語料庫在對外漢語教學中的應用〉，《廈門大學學報》哲社版，第4期，112-115。
- 北京語言學院語言教學研究所（1986），《現代漢語頻率詞典》。北京：北京語言學院出版社。
- 劉月華等（1983），《實用現代漢語語法》。北京：外語教學與研究出版社。
- 呂冀平、戴昭銘（1985），〈當前漢語規範工作中的幾個問題〉，《中國語文》，第2期，81-90。
- 呂叔湘（1979），《漢語語法分析問題》。北京：商務印書館。
- 呂文華（1999），〈短語詞的劃分在對外漢語教學中的意義〉，《語言教學與研究》，第3期，130-135。
- 齊滬揚（2000），《現代漢語短語》。上海：華東師大出版社。
- 任海波、王剛（2005），〈基於語料庫的現代漢語離合詞形式分析〉，《語言科學》，第4期，75-87。
- 朱德熙（1982），《語法講義》。北京：商務印書館。
- 李臨定（1990），《現代漢語動詞》。北京：中國社會科學出版社。
- 陸志韋（1957），《漢語的構詞法》。北京：科學出版社。
- 張壽康（1957），〈略論漢語構詞法〉，《中國語文》，第6期，1-8。
- 陳建民（1984），《漢語口語》。北京：北京出版社。
- 鄒立志、周建設（2007），〈動賓式動詞規範問題的深層理據〉，《語文研究》第1期，32-37。
- 余維欽、戴為華（1996），〈離合詞的特點及其辨認〉，《四川教育學院學報》，第12卷4，57-60。
- 國家漢辦（1992），《漢語水準辭彙與漢字等級大綱》。北京：北京語言學院出版社。
- 胡裕樹、範曉（1995），《動詞研究》。鄭州：河南大學出版社。
- 楊慶蕙主編（1995），《現代漢語離合詞用法詞典》。北京：北京師範大學出版社。
- 饒勤（1997），〈離合詞的結構特點和語用分析--兼論中高級對外漢語離合詞的教學〉，《漢語學習》，第1期，32-35。

趙元任（1968）、呂叔湘譯（1979），《漢語口語語法》中文本，北京：商務印書館。

趙淑華 張寶林（1996）、胡明揚主編，〈離合詞的確定及對其性質的再認識〉，《詞類問題考察》。北京：北京語言學院出版社。

[投稿：2009.1.6；審查：2009.2.16；接受：2009.03.16]

王海峰

Wang haifeng

中國北京北京大學對外漢語教育學院（100871）

School of Chinese as a Second Language, Peking University

china@pku.edu.cn

A Quantitative Study Based on the Large-Scale Corpus of the Chinese Separable Words

haifeng Wang

School of Chinese as a Second Language, Peking University

Based on the corpus-driven approach and the comprehensive study of the 207 Chinese separable words cognized in the Peking University corpus “the General Outline of Vocabulary for HSK”, this paper draws a conclusion that

- 1.The separation degree of the Chinese separable words is not as high as expected and the joint forms of the separable words are “regular” while the separated forms are “idiosyncratic”;
- 2.The separating forms can generally be reduced to 13 categories. Statistics shows that the most frequent separating form is “le”, secondly is complement, noun (pronoun), qualifier, and the least separating forms are verb components;
- 3.The separating particles and the numbers of the separable words with the separating forms form an inverse ratio: the fewer the particles of the separating forms are, the greater the number and higher the ratio the separable words have, and vice versa;
- 4.Most of the separable words with higher frequency are those used most frequently in daily life. The separable words with lowest frequency are those used least frequently in daily life.

Key words: corpus, separable words, quantitative, separated forms, frequency

