

《現代漢語新詞語資訊電子詞典》的研究與實現¹

Development and Study of the "Modern Chinese New Words Information Electronic Dictionary"

亢世勇*

Shiyong Kang

摘要

本文從四個方面介紹了我們正在開發中的《現代漢語新詞語資訊電子詞典》：
(1) 現代漢語新詞語的界定，(2) 新詞語詞典的開發思想，(3) 新詞語的採集與新詞語屬性資訊的描述，(4) 近四萬新詞語的歸類實踐。我們認定的新詞語是指 1978 年以來通過各種途徑產生的、具有基本詞彙沒有的新形式、新意義或新用法的語文詞語。除了詞形、詞義或用法任何一個方面“新”外，還要求必須是人們日常生活中普遍、廣泛使用的語文詞語，人名、地名以及專科術語都不屬於我們所說的“新詞語”。我們堅持開放的原則，儘量全面的採集收錄新詞語，用人機兩用的研究理念，以北京大學計算語言學研究所的《現代漢語語法資訊詞典》為模型打造一部收詞全面、資訊豐富、資源高度共用的現代漢語新詞語電子詞典，為新詞語的研究、中文資訊處理的研究提供一個寶貴的資源。目前已收錄新詞語近 4 萬，首先我們按照現代漢語詞類的“優勢語法”功能，給這四萬新詞語分類並歸類，然後，利用成熟的關聯資料庫（在 ACCESS 環境下實現）詳細地描述了每個詞語的屬性資訊。設立總庫一個，語法資訊庫三個，包括名詞庫、動詞庫、形容詞庫，另外還設立了構詞法庫，舊詞庫、外來詞庫、簡略詞庫。總庫和其他各庫通過“詞語、拼音、義項”三個欄位聯繫起來，構成了一個具有上下位關係的有機系統，便於資訊的提取。這些庫總共設立屬性欄位 200 多個，包括每個詞語的語音資訊、語義資訊、來源

¹ 本項研究得到中國國家哲學社會科學規劃專案(01CYY002)支援；

本文於 2002 年 4 月在臺北舉行的“第三屆中文辭彙語義學會議”上宣讀，會後根據專家的意見作了修改，謹致謝忱。

* 山東煙臺師範學院中文系(264025) E-mail:kangsy46@sohu.com Tel:0535-6672439
Shandong of China: Yantai Normal College - Chinese Language Dept. (264025)

資訊、構詞法資訊、句法資訊和部分語用資訊。本詞典是目前國內收詞量最大、描寫資訊最多的一部新詞語詞典。

關鍵字：中文資訊處理 新詞語 電子詞典

Abstract

We introduce the development of the Electronic Lexicon of Contemporary Newborn Chinese Words: (1) the definition of a newborn word, (2) the main principle behind constructing the lexicon, (3) the collection of newborn words and their feature descriptions of them, and (4) the classification of 40,000 newborn words. In our opinion, a new bornword is a character string that appeared after 1978 in a new form, with a new meaning and with a new usage. In addition, it must be frequently used and accepted, but the names of men and places are not newborn words according to our definition. The approach to collecting newborn words is quite unrestricted, that is, the more the better. Based on the Contemporary Chinese Grammatical Knowledge Base of the Institute of Computational Linguistics at Peking University, we have finished compiling a lexicon of almost 40,000 newborn words semi-automatically. The lexicon, we believe, is a worthy resource for research on Chinese word-building rules and Natural Language Processing. Firstly, classification is done based on the preponderant grammatical characteristics of each word, and then the detailed features are described in the database of ACCESS. The lexicon contains a total base and three grammatical bases (i.e., a noun base, verb base and adjective base); what's more, it also has an old word base, a loanword base and a acronym base. The entire base is related to the sub-bases through the fields of word, phonetic notation and semantics fields, which form a hypernymy hierarchy that is quite convenient for searching. Totally, there are more than 200 fields in the bases that give information regarding phonetic notation, semantics, sources, word building, syntax and pragmatics. Without doubt, this lexicon is one of the largest domestic lexicons available with the most detailed descriptions of newborn Chinese words.

Keywords: Chinese information processing, New words, Electronic dictionary

1. 引言

2001年我們獲得了中國國家社科規劃專案“《現代漢語新詞語資訊電子詞典》的開發與應用”（專案編號：01CYY002）。一年來，我們已按照規劃做了大量的工作，專案進展順利。本文從四個方面介紹《現代漢語新詞語資訊電子詞典》（以下簡稱“新詞語詞

典”) 的基本情況：(1) 現代漢語新詞語的界定 (2) 新詞語詞典的開發思想 (3) 新詞語的採集與新詞語詞典所描述的屬性資訊 (4) 近四萬詞語的歸類實踐

2. 現代漢語新詞語的界定

對於“新詞語”目前學術界有不同的看法，在全面考察了近 4 萬個新詞語並且借鑒、吸收了學術界新詞語研究成果的基礎上，我們認為新詞語可以定義為：通過各種途徑產生的、具有基本詞彙沒有的新形式、新意義或新用法的語文詞語。新詞語的特點在於“新”，“新”具體表現在詞形、詞義和詞語的用法上。鑒定新詞語的參照系是現代漢語基本詞彙的詞形、詞義和用法。只要在這三個方面的任何一點上與現代漢語基本詞彙不同，我們就認為它是新詞語。基本詞彙的代表是北京商務印書館出版的 96 版的《現代漢語詞典》和《漢語大詞典》。“新”還有時間的限定，即 1978 年以來出現的新詞語。我們認定的新詞語既有“新”的特點，同時強調了新詞語的使用範圍，即必須是在社會生活中廣泛使用的語文性質的新詞語，可以進入普通辭彙的新詞語，那些新出現的專業術語沒有增加新的普通辭彙意義的，不在我們認定的新詞語範圍內。我們認定的新詞語具體如下：

- (1) 新造詞語。比如“打假、扶貧、股盲、展銷、股市、高開、低走、哇噻、彩票、足彩、辣妹、酷裝、新新人類、哈韓族、哈日族、知本家、黑哨”等等。
- (2) 舊詞新用。這類詞語詞形是原有的，“新”主要表現在產生了新意義或有了新的運用。具體分為三種情況：A、原有的詞語增加了新的意義，如“下課、上課、氣候、跳槽、起飛、紅娘、窗口、下崗、亮相、新登場、跟進、充電、輸血、造血”等；B、原有的詞語有了新的用法。比如“結構”本來是名詞，但用為動詞，如：你為我結構人生；“運氣”原為名詞，用為形容詞，如：你這人很運氣。“火”原為名詞，用為形容詞，形容事物或人有聲勢，受歡迎。如：組織者們真沒想到晚會竟然這麼“火”。C、原有的詞語很長一段時間不用，又重新啟用，比如：“高就、賞光、黑道、綁票、撕票、夜總會、小姐、太太、金婚、銀婚”等。其中有些意義也發生了一些變化，比如“高就、賞光、太太、小姐”等原來主要用於地位比較高的人，有特指性，現在已經泛化，不論地位高低都可以用，變成了一種普通的說法。
- (3) 方言辭彙進入普通話辭彙。如“炒魷魚、發燒友、埋單、的士、連鎖店、服裝城、跳樓價、大出血、娛樂圈、拍拖、三級片、主打、金曲、勁歌、勁舞、搞笑、爽、靚、馬子、二奶、套磁、磁實、貓膩、腕兒、搓、傍大款、侃大山、膀爺”等。
- (4) 外來詞，從外族語借來的詞，又有：A、音譯詞如“的士、巴士、歐佩克、可口可樂、丁克、克隆、基因、託福、卡拉 OK、拜拜、酷(cool)、蔻(cute)、秀(show)、脫口秀(talk show)、血拼(shopping)、派對(party)、伊妹兒(E-mail)”

等；B、意譯詞，如“熱點（hot spot）、音樂電視（music television）、熱狗（hot dog）、超級市場（supermarket）”；C、音譯兼意譯詞，如“鏽射、呼啦圈、桑拿浴、迷你裙、吧女、酒吧、”等；D、直接使用日語的詞語，如：“放送、慰安婦、物語、寫真、人氣”等。

- (5) 簡略詞，在原有詞語的基礎上縮略而成的詞語。分為三種情況：A、簡稱詞，如“博導（博士研究生導師）、澳網（澳大利亞網球公開賽）、超市（超級市場）”；B、略語詞，如“嚴打（嚴厲打擊犯罪活動）、打假（打擊假冒偽劣商品）、防偽（防止假冒偽劣產品）、台資（臺灣人投入的資本）”；C、縮語詞，如“三講、三個代表、三假、三陪、三金”等。
- (6) 修辭用法穩定下來構成的新詞語。主要有：A、比喻引申，如“豆腐渣工程、枕頭風、撒胡椒麵、下毛毛雨、泡沫經濟、朝陽產業、白色消費、下海、撈人”等；B、借代，如：“菜籃子工程、白髮世界、白條案、老人頭”等。C、仿擬，比如：“煙民、股民、彩民、網民”，“空姐、海姐、吧姐、呼姐、網姐、空嫂、海嫂、吧娘、呼嫂”，“文盲、科盲、股盲、舞盲、網盲”、“網民、網友、網哥、網姐、網迷、網蟲、網蠅”等等。
- (7) 專用術語意義泛化、轉移，擴大使用範圍，轉為普通辭彙。如“軟體、硬體、啓動、熱處理、冷處理、黃牌、主旋律、套牢、觸電、放電”等。
- (8) 字母詞。主要有三類：A、純粹的字母詞，整個詞由英文字母構成，如“CT、IBM、CIA、TOFEL、GRE、CEO、ATM、CFO、BBS、CVD、DVD、VS、IT、IN、Q、VIP”等等；B、字母和漢字的組合，如“BP機、BP族、CALL機、E時代、E人類、IT界、IT業、夠IN、VIP卡、很Q”等等；C、數位和字母的組合，如“3D、3C、3S”等等。

3. 《現代漢語新詞語資訊電子詞典》的開發思想

3.1 新詞語研究的局限

現代漢語新詞語的研究受到了國內外的廣泛關注，學者們也做了大量的研究，產生了一些引人注目的研究成果。出版了新詞語詞典及詞語集三十多種、新詞語研究專著兩本，但是這些著作對新詞語的研究都有一定的局限。主要表現在以下方面：（1）這些研究成果都是印刷品，沒有有效的電子版成果，不能實現資源高度共用。（2）這些成果都是為人用的，而沒有考慮到機器使用，應用範圍受到了限制。（3）由於受到研究技術和研究條件的限制，各種詞典收詞量有限，詞語的解釋及引例都有欠妥之處，更重要的是詞典提供的信息量極其有限。由於以上的不足，造成現有的各種新詞詞典應用價值不高。

3.2 《現代漢語新詞語資訊電子詞典》開發的目標

- (1) 希望創建現代漢語新詞語研究的基礎平臺，實現資源高度共用，獲得較高的應用價值。本項研究利用電腦資料庫技術和相關的語料庫技術進行現代漢語新詞語的跟蹤研究，研究成果形式為有效、實用的電腦資料庫軟體，其中包括新詞語電子詞典和大規模的相關語料，這樣可以實現資源的高度共用，使其具有較高的應用價值。
- (2) 希望在漢語研究和中文資訊處理研究方面做出積極的貢獻。以往漢語的研究的資料和手段限制了漢語大規模的實用化的研究，由此造成的直接後果是嚴重制約了中文資訊處理的發展。本項研究利用電腦技術進行，積累了大量的機器可讀文件，為大規模的實用的漢語研究奠定了基礎，其研究成果——新詞語屬性資訊電子詞典以及新詞語的構詞規律可以直接應用於中文資訊處理的未登錄詞語識別，有利於提高中文資訊處理技術的水平。

3.3 《現代漢語新詞語電子詞典》的開發具體思路

介於目前有關新詞語的研究比較零散，而且新詞語的研究又有十分重要的作用，我們擬對新詞語進行大規模的比較完備的研究。具體思路為：

- (1) 儘量窮盡地收集現有的新詞語，做到全面、準確。目前已收錄新詞語近4萬，收錄了我們所能見到的所有新詞語。
- (2) 按照人機兩用的研究理念，打造一部適合於“人讀”和“機讀”的電子詞典。增加詞典的信息量，擴大詞典的使用範圍，提高其應用價值。
- (3) 以北京大學計算語言學研究所的《現代漢語語法資訊詞典》為模型，採用分類與屬性描述相結合的方法，在粗分詞類的基礎上對每個詞語語法語義屬性資訊進行詳細描述。具體採用成熟的關聯資料庫形式描述詞語和語法、語義屬性的二維關係，成果為資料庫文件格式的電子詞典。
- (4) 一部開放的詞典。本詞典在新詞語的收集及屬性的描述方面均堅持開放的原則，將跟蹤漢語辭彙的發展變化和漢語資訊處理的發展，不斷地收集、增加新詞語，增加新詞語屬性資訊的描述，以滿足實際需要。

4. 《現代漢語新詞語資訊電子詞典》詞語的採集與所描述的屬性資訊

4.1 新詞語的採集

首先利用我們自己開發好的《新詞語詞典資訊庫》和語料庫整理出一個新詞語詞表，然後按照我們的收詞原則——全面性原則、規範性與描寫性相結合原則、必要性原則、普

遍性原則、穩定性原則、音節原則等，從詞表中遴選出新詞語 3 萬多個，形成了新詞語詞典的基礎。此後，我們利用語言資訊處理技術不斷地從網上抓取新詞語及相關的例句集，不斷地擴充新詞語詞典。確定新詞語詞典中的詞目後，利用新詞語詞典資訊庫和包含《人民日報》1978 年以來的語料、《南方周末》創刊以來的語料以及人民日報報系其他報紙、人民網、光明日報、新民晚報等近年來語料的超大規模語料庫建立包含該詞語的例句集，考察這些詞語的意義和用法，描述其義項、語法屬性、語義屬性以及其他資訊等，從而開發出《現代漢語新詞語資訊電子詞典》。這些工作很大程度上利用電腦語料庫管理技術，在大規模機讀語料庫的支援下進行，能夠比較全面地考察每個新詞語的分佈環境，提高新詞語採集、收錄的合理性和資訊描述的準確度和覆蓋範圍，從而提升詞典的質量。

4.2 新詞語詞典屬性資訊的確立

新詞語詞典開發主要是為了學習、研究新詞語，特別是為中文資訊處理提供一個基本資源。為了達到這一目的，新詞語詞典屬性資訊包括了語音資訊、來源資訊、語法資訊和部分語義、語用資訊，涉及了新詞語形、音、義以及用法的主要方面。

新詞語詞典描述的主要屬性資訊包括以下方面：

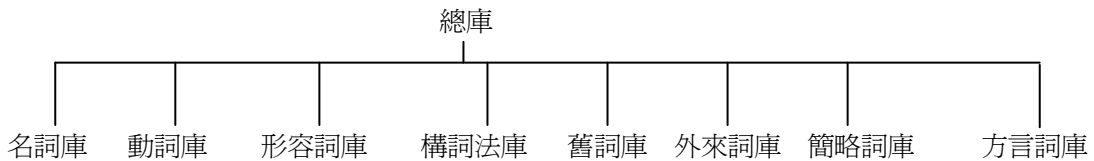
- (1) 詞的常規資訊。包括詞的讀音、義項、音節、例句等。
- (2) 語法資訊。按照北京大學計算語言學研究所的《現代漢語語法資訊詞典》的規格描寫新詞語的語法資訊。詞類體系沿用《現代漢語語法資訊詞典》的 18 個基本類，再加上成語、慣用語。詞類標記與其相同。各類詞語法屬性的設立在《現代漢語語法資訊詞典》基礎上有所改動，使其更加優化。
- (3) 構詞法資訊。構詞法主要分為單純構詞法和合成構詞法兩類。單純構詞法又分為單音單純詞、多音單純詞。多音單純詞又分為聯綿詞、音譯詞和疊音詞等。聯綿詞又分為雙聲、疊韻其他等。合成詞又分為複合式、重疊式、附加式三類。複合式又分為聯合式、偏正式、補充式、動賓式和主謂式等。附加式又分為兩種類型：“字首+詞根”、“詞根+尾碼”等。對於複合詞將構成複合詞的幾部分分解開來，分別標上該語素所屬的“詞性”，以便進一步考察由語素按照一定的構詞方法構成的新詞語的詞性的規律。
- (4) 產生途徑。根據我們的考察主要包括：新造詞，舊詞新用，方言詞進入普通話的辭彙，外來詞，簡略詞，修辭用法穩定下來構成新詞，術語擴大使用範圍產生新義。
- (5) 應用領域。應用領域的劃分是一個比較棘手的問題，我們大體上分為政治、經濟、法律、軍事、文化、科技、教育、衛生、體育、商業、工業、農業、生活、通用等，暫時作為工作規範，以後再逐漸調整。

- (6) 來源資訊，即該詞從那本詞典或哪些語料中來，如果很多本詞典都收錄了該詞，則說明該詞的複現率比較高，新詞語的身份更加確定。
- (7) 時間資訊。該詞語大致產生的時間，以詞典的引例時間為準；使用時間，以詞典出版時間為準。

4.3 新詞語詞典的結構與各個庫的主要屬性資訊

4.3.1 新詞語詞典的總體結構

新詞語詞典採用成熟的關聯資料庫技術（在 access 軟體下實現）。填入的資訊儘量以直觀明瞭的漢字、字母、數位表示。根據新詞語屬性的確立，資訊庫總體上包括三個方面五個庫。總庫一個，語法資訊庫三個（名詞庫、動詞庫、形容詞庫），構詞法庫一個。另外還設立了舊詞庫、外來詞庫、簡略詞庫和方言詞庫，對新詞語當中的舊詞新用、外來詞、簡略詞、方言詞等的有關資訊進行了描述。這幾個庫通過“詞語、拼音、義項”三個欄位連接。構成一個上下連接的有機系統，便於資訊的提取。新詞語詞典的總體結構如下：



4.3.2 各個庫所描述的主要屬性資訊

總庫主要描述的資訊有：詞語、拼音、義項、詞性、音節、產生途徑、領域、時間、來源等。

其他各庫共有的屬性資訊有“詞語、拼音、義項”，均從總庫中繼承，其他屬性資訊如下。

語法資訊庫中名詞庫主要描述了與名詞搭配的各種量詞，名詞的子類，能不能直接受數詞、數量詞、其他名詞、動詞的修飾，能受哪些代詞直接或加“的”後修飾，前接或後接成分，能不能作定語、主語、賓語，能不能直接或加“地”作狀語，以及能不能重疊、臨時充當量詞等。動詞庫主要描述的資訊有：動詞的子類——係詞、助動詞、趨向動詞、補助動詞、形式動詞、自主動詞、非自主動詞、內外動詞、存現動詞、離合詞等；構成的句式——“把”字句、“被”字句、兼語句、雙賓句、存現句等；充當的成分——定語、名詞性結構的中心語、單作謂語、賓語、狀語；後帶的成分——體謂准賓語、動時量補語、結果補語、趨向補語等；動詞自身形態的變化——前受“不、沒、很、正”的修飾、後跟“著、了、過”、VV、AABB、V—V、V了V、V了一V、VVO等。賓語、結果補語、趨向補語的詳細資訊將另行描述。形容詞庫描述的主要資訊有：子類、

直接作定語或加“的”後作定語、作謂語、補語、狀語或加“地”後作狀語或再加“很”後作狀語、作準謂賓、“有”的賓語、名詞性結構的中心語、AA 重疊及重疊後的詞性、ABAB、A 裏 AB、帶“著了過”、准賓語、趨向補語等。

構詞法庫描述的主要資訊有：1、構詞部件，分為“成分 1”“成分 2”“成分 3”，分別填入構成該詞語的成分的類別，其中有的是語素、有的是詞。2、構詞法，考察該詞語的構詞方式，主要分為：主謂、動賓、狀中、定中、補充、聯合、加字首、加尾碼等。3、詞性，填入詞語的詞性，從總庫中繼承來。4、音節，填入該詞語的音節數，從總庫中繼承來。

舊詞庫描述的主要資訊有：1、舊義，填入該詞語原來的意義；2、新義，填入該詞語新意義或新用法；3、詞性，填入該新詞語的詞性，如果詞性與原詞語詞性一致，則標詞性標記；如果改變了詞性則特別標明，如不及物動詞變為及物動詞，標為：Vt；及物動詞變為不及物動詞，標為：Vi。4、詞義演變途徑，考察由舊詞語演變為新詞語的詞義的演變途徑，主要有 36 類：（1）同用相比，（2）同果相喻，（3）同質相喻，（4）同狀相喻，（5）特定代普通，（6）具體到抽象，（7）同位相喻，（8）語素換義，（9）泛化，（10）個體代全體，（11）普通代特定，（12）使動化，（13）同感引申，（14）物件更換，（15）客體更換，（16）指稱物件擴大，（17）工具帶本體，（18）特徵、標誌代本體，（19）專化，（20）同形相喻，（21）以果代因，（22）同所相喻，（23）部分代全體，（24）動靜引申，（25）所在代（比如：山頭、大哥大），（26）主體擴大，（27）功用代本體，（28）語用（小姐），（29）社會原因（草業），（30）時空引申，（31）正反引申，（32）句法影響，（33）抽象到具體，（34）特指化，（35）本體代特徵，（36）現象代本體。（根據羅正堅的《漢語詞義引申導論》和徐國慶的《現代漢語辭彙系統論》歸納出來）5、演變類型，詞義演變的類型，主要有 9 類：（1）轉移，（2）擴大，（3）虛化，（4）轉類，（5）縮小，（6）貶降，（7）揚升，（8）弱化，（9）深化。

外來詞詞庫描述的主要資訊有：1、途徑，外來詞進入漢語的主要途徑，主要有：（1）音譯，（2）諧音，（3）音譯加漢語語素，（4）音兼意譯，（5）按照外語詞語的意義創造一個漢語詞語；2、語音變化資訊，主要有：（1）音素的替換，（2）音節的增減；3、意義變化，主要有（1）擴大，（2）縮小，（3）轉移，（4）保持原意；4、縮略，考察外來詞語是否有縮略；5、應用領域。

簡略詞詞庫描述的主要資訊有：1、原詞語，填入簡略詞的原型。2、簡略的類型：（1）簡稱，（2）縮語，（3）略語，（4）准縮略語。3、構成方式：將原詞語劃段，根據實際情況描述如何進行縮略的，比如“北京大學”——“北大”，其構成方式描寫為“a1b3”。4、同形：如有同形詞語，則有幾個填相應的數位。5、縮略方式：（1）縮合，如：奧林匹克運動會——“奧運會”；（2）節縮，如：電視連續劇——“連續劇”；（3）提取，如：中國高技術研究發展計劃綱要——“863 計劃”（該計劃的提出是 1986 年 3 月）；（4）其他，包括：A、用同義、近義詞語替換，如：浮式起重機——“浮吊”；

B、用上位詞語代替下位詞語，如：中華人民共和國教育委員會——“國家教委”；C、用英文中的字母縮略，如“MTV”。

舊詞新用庫、外來詞庫、簡略詞庫的開發主要是爲了研究新詞語的產生途徑以及主要原因。

各個庫屬性資訊的具體描述方法，請參閱拙作《〈現代漢語新詞語資訊（電子）詞典〉的開發應用》、《〈現代漢語新詞語資訊（電子）詞典〉的結構》。

5. 近四萬詞語的歸類實踐

進行語法資訊的描述首先要對新詞語進行分類和歸類。新詞語詞典所堅持的語法理論及詞類體系繼承了北京大學計算語言學研究所《現代漢語語法資訊詞典》所堅持的語法理論和詞類體系——片語本位語法體系。針對漢語詞類多功能的特點，我們堅持以“優勢語法功能”作爲詞類劃分和詞語歸類的標準。爲了明確漢語詞類的優勢語法功能，我們以《現代漢語語法資訊詞典》爲基礎進行統計，總結出漢語詞類語法特徵的分佈狀況以及優勢語法功能。漢語詞類優勢語法功能如下：

名詞：能受數量詞修飾，能作主賓語。

時間詞：修飾名詞、直接修飾“指量名”結構、作介詞“在”的賓語。

處所詞：作“在”的賓語、直接修飾名詞構成定中結構、能用“這兒、哪兒、那兒”指代。

區別詞：能加“的”後或直接修飾名詞作定語。

動詞：能受“沒”或“不”修飾、能帶時態助詞“著、了、過”、能單獨作謂語。

形容詞：作謂語、受“很”“不”修飾、作定語。

狀態詞：不受“不”“很”修飾、加“的”後修飾名詞、加“的”後修飾“數量名”結構、帶“的”後作謂語。

副詞：作狀語而不作定語。

根據以上“優勢語法功能”，對新詞語進行了分類，並對四萬新詞語逐一進行了考察，歸了類。在歸類中，我們注意到：1、這些優勢語法功能不是“對內具有普遍性”，因此不能包打天下，只是具有相對的普遍性，還有一些例外，需要用其他特徵協助判斷。2、很多優勢語法功能也不是“對外具有排他性”，一些優勢功能也可能是兩類詞共有的，比如能受“不”的修飾是動詞和形容詞共同的特徵，而不能受“不”的修飾又是名詞、時間詞、處所詞、區別詞、狀態詞的共同特徵。這種情況提醒我們給新詞語分類或歸類不能按照單一標準，而要綜合運用多個標準。

按照這些標準給現有新詞語分類和歸類情況如下：

名詞有：愛蟲、愛嬌、愛意、安居工程、安樂死、按摩小姐、奧星、奧運戰略、八卦新聞、吧女、吧台、吧蠅、霸氣、白金唱片、白領、白領犯罪、白領麗人、白判、白色公害、白色收入、板寸、傍姐、保護傘、保健菜、波霸、長期飯票、超級恐龍、炒家、搞笑片、AA制、BBS、e-Book、SOHO族、等等 24874 個。

動詞有：挨宰、暗箱操作、拔份、把脈、罷網、白領化、擺平、扮靚、扮酷、扮靚、棒殺、傍、傍大款、包二奶、包二爺、包裝、煲電話粥、煲網、保級、暴跌、保廉、暴走族、曝光、爆炒、爆棚、蹦迪、蹦極、逼宮、飆、飆車、飆價、炒樓花、觸網、豐胸、搞掂、Call、等等 12006 個。

形容詞有：霸氣、暴露、倍兒棒、慘、火、火爆、酷、靚、帥、爽、靚麗、IN、Q、Cool 等等 1002 個。

區別詞有：非常、海量、候鳥型、綠色、新銳、主打、袋裝、獨資、兩栖、落地式、程式控制、等等 119 個。

副詞有：慘、好、絕、斃、倍兒、倍加、等等 43 個。

成語有：築巢引鳳、引咎辭職、招商引資、一網情深、優化組合、友情出演、友情客串、心靈雞湯、霧裏看花、閃亮登場、強強聯合、牽線搭橋、錢權交易、夢中情人、美麗凍人、快樂老家、高開低走 236 個，

慣用語有：愛誰誰、愛情走私、愛心大放送、把蛋糕做大、別理我，煩著呢、找不著北、有沒有搞錯、玩兒深沈、沒事偷著樂、老鼠愛大米、空手套白狼、跟著感覺走、第一次親密接觸、常回家看看、瀟灑走一回、玩的就是心跳、等等 799 個。

歎詞有：哇噻、yeah2 個。

參考文獻

- 亢世勇·《現代漢語新詞語資訊(電子)詞典》的開發應用，《辭書研究》，2001(2)：55—63。
- 亢世勇·《現代漢語新詞語資訊(電子)詞典》的結構，載於《資訊網路時代中日韓語文現代化國際研討會論文集》，香港：香港文化教育出版社，2000：276—281。
- 亢世勇·語料庫技術在新詞語詞典開發中的具體應用，載於《中國辭書論集 2000》，北京：中國大百科全書出版社，2001：291—300。
- 王鐵昆·新詞語的判定標準與新詞新語詞典編纂的原則，《語言文字應用》，1992(4)：14—20。
- 劉一玲·尋求新的色彩，尋求新的風格——新詞語產生的重要途徑，《語言文字應用》，1993(1)：85—90。
- 季恒銓·新詞新語詞典編纂的新收穫，《語言文字應用》，1993(1)：77—80。

- 語用所“新詞新語新用法研究”課題組·整理漢語新詞語的若干思考,《語言文字應用》, 1993(3):65—76。
- 張志毅、張慶雲·新時期新詞語的趨勢與選擇,《語文建設》,1997(3):15—18。
- 羅正堅·《漢語詞義引申導論》,南京:南京大學出版社,1996。
- 俞士汶·《現代漢語語法資訊詞典詳解》,北京:清華大學出版社,1998。
- 徐國慶·《現代漢語辭彙系統論》,北京:北京大學出版社,1999。
- 姚漢銘·《新詞語·社會·文化》,上海:上海辭書出版社,1999。
- 於根元·《網路語言概說》,北京:中國經濟出版社,2001。

