

Corpus in Foreign Language Teaching and Research

Le corpus et les recherches de l'enseignement-apprentissage des langues étrangères

語料庫與外語教學研究

Zhou Xiping

周新平

Received 26 March 2006 ; accepted 19 September 2006

Abstract Corpus-based language research has been long prospered since the middle of last century. Corpus is therefore frequently used in foreign language (mostly English) teaching and research due to the fundamental principles of modern Corpus Linguistics along with the colorful resources of word-banks and the corresponding tools, especially in western countries. In China, the related literature found its way from introducing the foreign researches to our own practice into this field. As a conclusion, corpus and Corpus Linguistics can be closely connected with and widely applied in foreign language teaching and research with a predictable bright future.

Keywords: corpus, Corpus Linguistics, foreign language teaching and research

Résumé Le moyen de recherches sur le corpus a connu un développement rapide depuis le milieu du siècle précédent et a atteint la maturité aujourd'hui. En raison de l'importance de la linguistique de corpus et de la méthode de recherches sur le corppus, et étant donné ses ressources riches ainsi que les facilités apportées par les outils de recherche, le corpus est appliquée amplement dans les recherches linguistiques notamment dans celles de l'anglais. A l'étranger, l'étude de la linguistique de corpus a débuté tôt et a donné beaucoup de fruits ; le travail du milieu des langues étrangères chinois dans ce domaine a commencé par la présentation du corpus étranger et sa situation d'étude, et puis procède à des applications pratiques. En somme, il existe des relations étroites et diverses entre le corpus et les recherches de l'enseignement-apprentissage des langues étrangères, et les recherches de l'enseignement-apprentissage des langues étrangères basant sur le corpus présente une bonne perspective.

Mots-clés: corpus, linguistique de corpus, recherches de l'enseignement-apprentissage des langues étrangères

摘要 語料庫研究方法自上個世紀中葉以來迅速發展，到今天已空前繁榮。由於語料庫語言學和語料庫研究方法的重要性，又由於豐富的語料庫資源及檢索工具提供了實踐的可能和便利，語料庫被廣泛應用於語言研究特別是英語教學研究中。國外語料庫語言學研究起步早，成果多；我國外語界在這方面的工作主要是從介紹國外語料庫及其研究現狀開始，然後進行一些應用實踐。總之，語料庫與外語教學研究有著緊密的、多層面的關係，基於語料庫的外語教學研究大有可為、前景廣闊。

關鍵詞: 語料庫；語料庫語言學；外語教學研究

1. 語料庫與語料庫語言學

語料庫 (corpus)是一個由大量在真實情況下

使用的語言資訊集成的專供研究使用的資料庫。語料庫語言學作為一種研究語言的方法，可追溯到上個世紀，甚至更為久遠。當時的研究手段還只停留在卡片製作和人工檢索的階段，如 19 世

紀和 20 世紀早期的兒童語言學的研究主要是以詳細記錄孩子的話語發展為基礎的。

現代語料庫語言學 (modern corpus linguistics) 是二十世紀中葉開始的語言研究科學。何安平 (1997) 指出, 將大量語料作電腦化處理, 即用電腦進行注釋加工、電腦解碼以及電腦檢索、取樣和統計分析等則是近二、三十年的事。^{1(P21)} Kennedy(1998)、Biber 等(1998)有專著對現代語料庫語言學的發展、研究現狀和研究成果作了詳盡論述,^{2,3}他們的專著是語料庫語言學的必讀書目。

語料庫研究方法自上個世紀中葉以來迅速發展,到今天已空前繁榮。Leech(1993)指出,“從科學方法的角度,語料庫研究方法是一種更為強有力的方法,因為其結果是可以驗證的。”^{4(P106)}據美國加州大學語言學家 Edwards 1993 年的不完全統計,80 年代以來建成並投入使用的各類語料庫達 50 多個,按語種分佈如下:英語 24 個、德語 7 個、法語 4 個、西班牙語、義大利語、芬蘭語、丹麥語、瑞典語各 2 個。南斯拉夫語、葡萄牙語、愛脫尼亞語等也都建立了語料庫。另外還有法語、日語、馬來語、漢語語料庫等正在努力建立之中。這不能不說明了語料庫語言學與語料庫研究方法的強大生命力。國外語料庫語言學研究起步早,成果多;與其他領域相比,國內這方面的研究稍顯薄弱,有待加強。特別是廣大外語教師應進一步瞭解國內外主要外語語料庫及其檢索工具的特點,並有意識地將它們應用於外語教學研究中。

2. 主要外語語料庫及語料庫檢索工具

2.1 國外主要英語語料庫^{5, 6, 7}

¹何安平. 談語料庫研究[J]. 外國語 1997 (5).

² Biber, D., S. Conrad and R. Reppen. Corpus Linguistics[M]. Cambridge: Cambridge University Press, 1998.

³ Kennedy, G. An Introduction to Corpus Linguistics[M]. Beijing: Foreign Language Teaching and Research Press, 1998.

⁴ Biber, D., S. Conrad and R. Reppen. Corpus Linguistics[M]. Cambridge: Cambridge University Press, 1998.

⁵戴煒棟,張愛玲. 語料庫、電腦、語言學[J]. 外國語, 1999 (6).

2.1.1 COBUILD 語料庫: 由 John Sinclair 主持完成, 它是目前為止最大的英語語料庫, 所含的語料已超過 5 億詞。COBUILD 是八十年代語料庫語言學復興的標誌, 它使人們認識到電腦語料庫的巨大潛在用途, 導致了建立和利用大型電腦語料庫的熱潮, 產生了豐碩成果。其中的英語庫 (The Bank of English), 收詞兩億五千多萬, 由各種類型的書面語資料和口語資料組成; 它是朗文出版公司出資、柯林斯出版公司與伯明罕大學聯合開發的; 其前身就是 80 年代 Sinclair 領導的英國伯明罕大學語言研究小組建立的 BCET (Birmingham Collection of English Tests) 語料庫, 當時容量為 730 萬詞次左右。

2.1.2 英國國家語料庫 (British National Corpus, 簡稱 BNC): 收詞一億多, 由牛津大學出版社、朗文出版公司、錢伯斯—哈洛普出版公司、牛津大學電腦服務中心、蘭喀斯特大學英語電腦研究中心以及大英圖書館聯合開發。

2.1.3 英語用法調查 (The Survey of English Usage, 簡稱 SEU): 收詞十萬, 60 年代以英國英語為物件而建立的語料庫, 其中約一半書面語, 一半口語。口語部分又與其他語料組成了 The London-Lund Corpus。

2.1.4 國際英語語料庫 (The International Corpus of English, 簡稱 ICE): 收詞一百萬, 口語與書面語各一半。由倫敦大學學院英語系 1988 年承建, 語料取自所有英語國家。

2.1.5 布朗語料庫 (The Brown Corpus): 收詞約十萬, 以 1961 年以後美國書面語為基本內容。

以上是國外幾個著名的英語語料庫, 其他常見的還有倫敦—倫特語料庫 (The London-Lund Corpus)、蘭卡斯特—奧斯陸/伯爾根語料庫 (The Lancaster-Oslo/Bergen Corpus, 簡稱 LOB)、瑪喀裏語料庫 (The Macquarie Corpus)、朗文—蘭喀斯特語料庫 (The Longman-Lancaster Corpus) 等。

2.2 我國已有的及建設中的主要英語語料庫:^{8, 9}

⁶王建新. 語料庫語言學發展史上的幾個重要階段 [J]. 外語教學與研究 1998(4).

⁷梁旭紅. 語料庫語言學研究綜述[J]. 晉東南師範專科學校學報, 2001(1).

⁸戴煒棟,張愛玲. 語料庫、電腦、語言學[J]. 外國語, 1999

早期有代表性的是上海交通大學的 JDEST 科技英語語料庫。1982 年由黃人傑、楊惠中主持完成，收詞 100 萬。由 2000 篇 10 個專業的科技英語文本組成，這些文本隨機取自課本、雜誌、摘要等，含美國英語和英國英語。

近期有代表性的是桂詩春和楊惠中教授主持的中國學習者英語語料庫。它是一個 100 萬詞的中國英語學習者書面語語料庫，而且進行了言語失誤標注。

我國其他譯建成及再建的英語語料庫還有香港科技大學的科技英語語料庫、江漢石油學院的石油英語語料庫、香港理工大學英語學習者語料庫、廣東外語外貿大學的中國學生交際英語語料庫、北京外國語大學的英語教學語料庫等。

2.3 我國已建成的主要日語語料庫：^{10, 11}

國內日語語料庫的建設相對較晚，因而成果也相對較少。目前只有北京日本學研究中心開發的兩個語料庫，一個是徐一平等在 1999 年創建的具有 2000 萬字的大規模的《中日對譯語料庫》，另一個是曹大鋒等 2003 年開發研製的日語教材語料庫，這是國內第一個日語精讀課教學研究資料庫。

2.4 主要語料庫檢索工具簡介^{12, 13}

2.4.1 Concordance 系列

英國伯明罕大學語言研究小組開發的，具有詞語檢索 (concordancing) 並共現上下文的功能，可對索引單位在語料庫中所有的相匹配字串迅速可靠地定位，並載出規定長度的上下文。John Sinclair(1991)在其《語料庫、檢索與搭配》一書中例舉了詞語檢索的作用。¹⁴

2.4.2 Wordsmith 系列套裝軟體

1996 年牛津大學出版社推出 Mike Scott 所編

(6).

⁹ 王建新. 介紹當代三個英語語料庫[J]. 外語教學與研究, 1996(3).

¹⁰ 徐一平. 中日對譯語料庫的研製與應用研究論文集[C]. 外語教學與研究出版社 2004 年;

¹¹ 曹大鋒. 日語教學與教材創新研究——日語專業基礎課程綜合研究[C]. 高等教育出版社 2006 年 10 月。

¹² 王建新. 語料庫語言學發展史上的幾個重要階段[J]. 外語教學與研究, 1998(4).

¹³ Sinclair, J. Corpus, Concordance, Collocation[M]. Oxford: Oxford University Press, 1991.

¹⁴ Sinclair, J. Corpus, Concordance, Collocation[M]. Oxford: Oxford University Press, 1991.

制的 Wordsmith 套裝軟體。這套軟體不僅提供了詞語檢索並共現上下文的常用功能，還可統計計算出索引單位搭配行為的 T(或 Z)值、資訊相關(即 MI: mutual information)值和 X^2 (chi-square)值，並分別予以列表顯示。它還能快速識別某種文體中經常出現的多詞集合，且提供在整個語料庫中對英語教師和英語學習者具有價值的語言特徵。

2.4.3 其他檢索工具

常見的檢索工具還有 KWIC(key word in context)、Wordcrunch 等。它們有的可在網上使用，有的索引軟體可進行布林組合式(Boolcan Combination)檢索，還有的可以檢索各種詞綴。有志於運用語料庫進行外語教學研究者應瞭解、掌握其中的一兩種。

3 語料庫在外語教學研究中的應用

由於語料庫語言學和語料庫研究方法的重要性，又由於豐富的語料庫資源及檢索工具提供了實踐的可能和便利，語料庫被廣泛應用於語言研究特別是英語教學研究中。我國外語界在這方面的工作主要是從介紹國外語料庫及其研究現狀開始，然後進行一些應用實踐。謝應光(1996)、何安平(1997)、郭傑克(1997)、顧曰國(1998)、馮躍進(1998)、王建新(1999)、廖七一(2000)、梁旭紅(2001)、施建軍(2003)、曹大鋒(2005)等曾撰文論述過語料庫在外語教學研究中的應用，概括起來大致有如下幾個方面：

3.1 語料庫與詞典編纂及辭彙研究

引述實例是詞典編纂的傳統。機讀語料庫為語言提供了豐富、快捷的語言材料，改變了詞典編纂者利用語料的方式。¹⁵ (P36) 這不僅使詞典的編纂和修訂進度能大大加快，而且能及時提供新的語言資訊，同時大量的自然語言實例會使詞義更加完整和確切。例如，以 COBUILD 為基礎已出版了二十幾種詞典、語法書、同義詞庫和英語教材。

3.2 語料庫與句法研究

¹⁵ 朱樂紅. 語料庫對語言研究及外語教學的作用[J]. 外語與外語教學, 2000(3).

句法研究成為目前利用語料庫開展研究最多的語言層面。其原因首先是因語料標和電腦檢索手段的發展,使得對語言中有代表性的語法現象加以量化分析更為容易;其次是通過量化具備代表性的語法現象來檢測根據語法理論所提出的假設。如 Romaine(1996)以 BNC 為依據,研究了雙重比較級現象。以 J. Aarts 為首的研究者,根據語言學家的內省及現有的語法描述設計了一套形式語法,將其輸入電腦在語料庫中運行,來檢測該語法對語料庫資料的解釋程度;然後根據語料分析結果修訂語法,通過這種研究方法他們認為語料庫會使理論得到核對總和進一步的發展。^{16 (P73)}最近,陳萬霞(2002)通過中國學習者語料庫對英語被動語態的習得情況進行了研究。¹⁷

3.3 語料庫與語義研究

語料庫可用來對為語義提供客觀標準,並適當考慮到語義的不確定性。^{18 (P5)}如 Mindt(1991)指出,詞項和語言結構的意義常常是根據語言學家本人的直覺來描寫的,而事實上語義區別是同句法、詞法和上下文的語篇相關;通過語料庫來調查這些相關成分,可以找到特定語義區別的客觀指示。^{19 (P36)}

3.4 語料庫與社會語言學

利用語料庫開展的社會語言學研究目前主要集中在語言與性別領域。其中運用最多的是英語中的性別歧視現象。Kjellmer 1986 年曾經調查了 man/men ,woman/women 及其它相應的代詞,結果表明男性代詞出現的頻率要比女性代詞出現的頻率高。^{20 (P5)}

3.5 語料庫與外語教學和測試

迅猛發展起來的語料庫語言學近年來出現了從純語言研究發展到語料庫與語言教學相結合研究的新趨勢。語料庫在語言教學方面得到了直接

或間接的應用。^{21 (P74)}首先,師生可以從語料庫中得到現實生活中真正的語言材料,以瞭解本族語的人們是如何表達某些意義的;其次,利用大型語料庫的詞頻統計結果能編制出科學、合理的教學詞表。如我國大學英語教學大綱的分級辭彙表就是依據 JDEST、Brown 及 American Heritage 等進行定量分析制定的。而 2001 年修訂發行的《高職高專英語課程教學基本要求》中的片語表則採用了大學專科英語基本片語研究專案(張淑標,黃開勝,2001)的成果,該專案運用英語語料庫對大量英語片語進行了定量分析,並根據使用頻率確定了專科學生應掌握的基本片語。^{22 (P84-86)}

作為我國首例教材語料庫,日語教材語料庫在對我國現行的教材進行多角度、多層面的深入研究的嘗試中起到了非常重要的作用,它能幫助日語教師全面把握現有教材的長處和短處,為新教材的開發提供了詳盡的資訊。此外,很多英語教師發現,測試中很多試題不能反映英語的真實面貌。而用語料庫原則編寫的試卷更真實、客觀,具有語用價值。在目前強調交際能力,注重語用的時代,語料庫中的實際例子要比語言學家、語法學家、詞典學家為解釋某一語法現象、某一詞語的含義依靠直覺編寫的例證來得更直接、更自然、更生動真實、更能表現出語言的自然本質。²³

3.6 語料庫與專門用途英語(ESP)

研究人員建立專門用途英語語料庫來提供特定領域的語言學習材料。如上海交大的 JDEST 語料庫、香港科技大學的科技英語語料庫、江漢石油學院的石油英語語料庫等。由於其自身特點 ESP 需要由語料庫語言學作為支持,以便深入地開展研究;而 ESP 分類語料庫又由於 ESP 研究者的要求而將被建設得更全面,自動化程度更高,漸趨完善;ESP 與語料庫建設相互作用、相互促進(陳明瑤,2000)。^{24 (P60-61)}賀燦文,周江林(2001)進行了英語科研論文中綜述性動詞的語料庫研究,從語義、語用的角度對綜述性動詞進行了專

¹⁶梁旭紅.語料庫語言學研究綜述[J].晉東南師範專科學校學報,2001(1).

¹⁷陳萬霞.從中國學習者語料庫看英語被動語態的習得[J].外語教學與研究,2002(3).

¹⁸戴煒棟,張愛玲.語料庫、電腦、語言學[J].外國語,1999(6).

¹⁹朱樂紅.語料庫對語言研究及外語教學的作用[J].外語與外語教學,2000(3).

²⁰戴煒棟,張愛玲.語料庫、電腦、語言學[J].外國語,1999(6).

²¹梁旭紅.語料庫語言學研究綜述[J].晉東南師範專科學校學報,2001(1).

²²黃開勝,張淑標.大學專科英語基本片語研究實驗報告[J].黃岡師範學院學報,2001(4).

²³史順良,馮躍進.語料庫在語言測試中的應用[J].山東外語教學,1997(4).

²⁴陳明瑤.ESP 與語料庫建設[J].外語研究,2000(2).

門探討。²⁵

3.7 語料庫與翻譯

語料庫語言學的發展對語言研究以及與語言研究密切相關的翻譯研究產生了很大影響。如廖七一(2000)在《語料庫與翻譯研究》一文中討論了與翻譯有關的三類語料庫(特別是英國翻譯英語語料庫)對翻譯教學、探索翻譯規範與驗證翻譯理論家提出的翻譯普遍性的積極意義。²⁶

馮躍進, 陳偉(1999)的漢語“副職”英譯的語料庫調查研究和陳偉(2000)的漢語“會議”英譯的語料庫調查研究分別在英語詞庫裏對英語中表示“副職”和“會議”的辭彙進行了詳盡的調查研究, 對其意義和語用域作了明確的區分, 並援引真實的例句對其搭配和語用範圍進行了闡釋和說明。^{27, 28}

徐一平研製的中日對譯語料庫收入了多種體裁的現代日語和漢語的真實文本及其對譯文本達2000萬字, 為中日語的語法, 句法的對比研究提供了極大的方便。

此外, 語料庫與文體學、心理語言學、語用學和語篇分析等學科和領域也存在著直接或間接的聯繫。可見, 語料庫與外語教學研究有著緊密的、多層面的關係, 基於語料庫的外語教學研究大有可為、前景廣闊。

4 語料庫及其相關研究的發展前景

對於語料庫語言學的發展前景, J.Svartvik(1992)預測, “電腦將運行更快、體積更小、價格更低; 語料庫將規模更大、品質更好、利用率更高”。²⁹ (P74)

馮躍進, 潘番(1998)指出: 現在的語料庫大多只是文本語料, 而未來的語料庫則將把聲音、圖

像等媒體資訊都聚錄其中, 形成文本型、聲音型、圖像型和混合型語料庫。如英國伯明罕大學語言研究小組正與英國廣播公司合作建立 BBC 英語聲像語料庫, 以期更深入地進行英語音韻學、文化學等方面的研究。可以相信, 多媒體語料庫將為語言研究提供更多方便, 為語言學習創造更佳的學習條件。在語料庫研究領域裏, 可望在不久的將來看到百花齊放的景象, 種類繁多的語料庫將會揭開語言研究史上新的一頁。³⁰ (P10)

戴煒棟, 張愛玲(1999)認為以下三個趨勢是可以預見的: 第一, 使用的人會遍及語言研究、描述或教學的所有領域; 第二, 語料庫收集的內容會越來越廣泛; 第三, 可以隨身攜帶的分析工具會使相關研究越來越方便。³¹ (P7)

筆者在這裏擬補充三點: 第一, 越來越多的語料庫會逐漸網路化; 第二, 獲得和使用各種語料庫會越來越便利和便宜; 第三, 我國英語學習者語料庫、仲介語料庫及外漢雙語語料庫的開發應用會越來越發達。

參考文獻

- [1] Leech, G. The State of the Art in Corpus Linguistics. In Karin Aijmer & Bengt Allenberg (eds.). *English Corpus Linguistics. London and New York: Longman*, 1993.
- [2] 謝應光. 語料庫語言學與外語教學[J]. *外語教學與研究* 1996, 3.
- [3] 顧曰國. 語料庫與語言研究[J]. *現代語言學*, 1998, 1.
- [4] 郭傑克. 語料庫、語料庫語言學與英語教學[J]. *華南理工大學學報* (自科版) 1997(增刊).
- [5] 王建新. 我國在語料庫語言學研究方面的部分進展(概述) [J]. *外語與外語教學* 1999, 3.
- [6] 施建軍, 徐一平. 語料庫與日語研究[J]. *日語學習與研究*, 2003年(4).

作者簡介: Zhou Xinping(周新平), Zhongnan University of Economics and Law, Wuhan, Hubei, 430073, P.R. of China. 中國中南財經政法大學外國語學院。

通信地址: 周新平, 中南財經政法大學外國語學院, 中國湖北武漢, 430073。

E-mail: joanzhou2@163.com

²⁵ 賀燦文, 周江林. 英語科研論文中綜述性動詞的語料庫研究[J]. *外語學刊*, 2001(4).

²⁶ 廖七一. 語料庫與翻譯研究[J]. *外語教學與研究*, 2000(5).

²⁷ 馮躍進, 陳偉. 漢語“副職”英譯的語料庫調查研究[J]. *外國語*, 1999(2).

²⁸ 陳偉. 漢語“會議”英譯的語料庫調查研究[J]. *上海科技翻譯*, 2000(2).

²⁹ 梁旭紅. 語料庫語言學研究綜述[J]. *晉東南師範專科學校學報*, 2001(1).

³⁰ 馮躍進, 潘番. 語料庫語言學的最新動態及未來發展趨勢[J]. *山東外語教學*, 1998(4).

³¹ 戴煒棟, 張愛玲. 語料庫、電腦、語言學[J]. *外國語*, 1999(6).