

不同母語背景華語學習者的用詞特徵：  
以語料庫為本的研究\*

**Salient Linguistic Features of Chinese Learners with  
Different L1s: A Corpus-based Study**

張莉萍<sup>+</sup>

**Li-ping Chang**

**摘要**

本文採用語料庫語言學方法，對比不同母語背景學習者語料庫，試圖尋找不同母語背景學習者的詞語特徵。首先透過主題關鍵性方法快速找出不尋常高頻、不尋常低頻關鍵詞語，進一步從語言類型、詞類概念、文化因素等面向來解釋這些關鍵詞語突出的原因。這些特徵包括母語為英語學習者代名詞和能願動詞「會、可以」的使用、相對少用句尾語氣助詞、多用量詞「個」；日、韓學習者在假設句使用上優先選用「...的話」形式、因果句優先選用「所以」等等。

**關鍵詞：**華語、學習者語料庫、主題關鍵性、中介語對比、語言類型、母語遷移

**Abstract**

The study aims to explore the salient linguistic features of Chinese lexical items from different L1s learners. The research method is corpus-based, including comparing the learner corpus and the native-speaker corpus, as well as sub-corpora for different L1s. The learner corpus which consists of more than 1.14 million Chinese words from novice proficiency to advanced learners' texts is mainly from the computer-based writing Test of Chinese as a Foreign Language (TOCFL). The

---

\*本研究得到科技部計畫(NSC 101-2631-S-003 -003)、科技部跨國頂尖研究中心計畫(NSC 103-2911-I-003-301)、教育部邁向頂尖大學計畫以及國立台灣師範大學華語文與科技研究中心部分經費補助，特此感謝。並感謝本期刊兩位匿名審查人給予之寶貴意見。

<sup>+</sup> 國立台灣師範大學國語教學中心, Mandarin Training Center, National Taiwan Normal University  
E-mail: lchang@ntnu.edu.tw

sub-corpora of Japanese, English, Korean, Vietnamese, Indonesia and Thai are observed. Japanese corpus is top 1, which occupies twenty four percent of the total data, followed by English, Korean, and etc. And the native corpus is from the Academia Sinica balanced corpus. Through the overuse or underuse linguistic forms and keyword-keyness analysis, some salient features are discovered. For examples, comparative to Chinese learners with other L1s, English language background learners show the unusual high frequency on pronouns and unusual low frequency on sentential final particles in Chinese writing. And Japanese as well as Korean background learners tend to overuse the post form ‘*de hua*’ instead of ‘*ruguo*’ when expressing the ‘*if*’ sentence, and overuse ‘*suoyi*’ instead of ‘*yinwei*’ when expressing the cause-effect relation. The article also provides possible explanations for these results from the aspects of learners’ native language typology, linguistic structure, syntactic category and culture.

**Keywords :** Mandarin Chinese, Learner Corpus, Contrastive Inter-language Analysis , Keyword-keyness, CEFR, Language Transfer

## 1. 前言

語料庫語言學興起於 20 世紀 80 年代，藉由語料的大量蒐集及標記工作，提供客觀、真實的語料給語言學研究者分析以及做為教學用例。同時間，第二語言學習者或外語學習者產出的語料也受到重視，所建置的語料庫稱為學習者語料庫（learner corpora）或中介語語料庫（interlanguage corpora）。最早的學習者語料庫是 1980 年代晚期由朗文出版集團所建立的朗文學習者語料庫（Longman Learners’ Corpus），約一千萬詞規模，語料來源為各國英語教師所提供的學生作文或考試語料。之後所建立的學習者語料庫多為語料加註了偏誤標記（Díaz-Negrillo & Fernández-Domínguez, 2006），並為不同母語背景的學習者建立子語料庫，例如，ICLE (International Corpus of Learner English) 為英語學習者建立了 14 個子語料庫，目前子庫還在不斷增加中。學習者語料庫可以讓語言研究者或教學者觀察學習者的實際運用情況，對學習者的語言特徵和語言發展進行全面而有系統的描述和對比分析<sup>1</sup>（Granger, 1998; Granger, Hung & Petch-Tyson, 2002; Hawkins & Buttery, 2009）。

關於漢語學習者語料庫的建置相對晚些，以 2009 年在大陸正式公開的「HSK 動態作文語料庫」為代表<sup>2</sup>，該語料庫是母語非漢語的外國人參加高等漢語水平考試（HSK）紙筆作文考試的語料，於 2003 年開始建置，蒐集了 1992-2006 年的部分外國考生的作文答卷，共計 11569 篇，424 萬字。雖然有四百多萬字的規模，但僅限當初參加 HSK 高

<sup>1</sup> 本文使用「特徵」一詞，是指廣義的語言使用特徵，只要是足以描繪出學習者使用語言的個別特色，都可稱為特徵，非指理論語言學中的抽象特徵。

<sup>2</sup> 可參見 <http://202.112.195.192:8060/hsk/login.asp>

等考試的文本<sup>3</sup>，也就是這些語料多是高等程度學習者的寫作語料。台灣的「TOCFL 學習者語料庫」也是類似性質，蒐集 2006 年至今參加華語文能力測驗（TOCFL）考生的作文，不同的是，這個測驗是電腦考試，也就是考生直接於線上輸入文字。現階段約 114 萬詞（174 萬字左右），語料涵蓋 42 種不同母語背景、不同能力考生所寫的作文，共 5092 篇，130 個主題，語料量仍在持續擴充中，關於語料庫的建置與語料分布請參考張莉萍（2013a）。

在對外漢語方面，以語料庫為本的語言特徵研究相對地少，在詞彙特徵方面，張莉萍（2012）利用語料庫文本覆蓋率的概念，驗證能力越高的學習者所需要的詞彙量越大，並從教學與學習的角度，建議對應於 CEFR（Common European Framework of Reference for Languages）（COE, 2001）不同等級學習者所需要的詞彙量，A2 為 1000 個詞、B1 為 2300-3000 個詞、B2 為 4500-5000 個詞、C 級為 8000-10000 個詞<sup>4</sup>。張莉萍（2013a）也嘗試分析了不同能力學習者使用高頻動詞的情況，與英語學習一樣，呈現同樣的趨勢（Hawkins & Buttery, 2009:164），即，隨著學習者語言水平的提高，使用高頻動詞的頻率越趨近本國人（p.147）。在語法特徵方面，張莉萍（2013b）則是利用本國人語料庫與學習者語料庫的對比，企圖找出不同能力階段學習者的句式表現特徵，例如，把字句、被字句。這些研究結果可以直接或間接應用在教學與評量上，本研究的目標則是企圖找出不同母語背景學習者的用詞特徵。

## 2. 研究方法與步驟

這個研究採用中介語對比分析方法（contrastive interlanguage analysis, 以下簡稱 CIA），CIA 是語料庫語言學與第二語言習得研究結合後的新領域，主要是針對兩類語料進行對比，一是本國人語料庫和學習者語料庫的比較；一是不同母語背景學習者語料庫之間的比較（Granger, 1998: 12）。應用前者對比分析可以找出學習者少用（underuse）或過度使用（overuse）的語言特徵；應用後者對比分析，則可能找出針對不同母語背景學習者的語言特徵。本研究著重在後者的描述與討論，所選用的學習者語料庫即是前述的 TOCFL 語料庫，主要是這個語料庫建置時，即載有學習者的母語背景資料；HSK 則僅有國籍資料。

為了觀察不同母語背景學習者的用詞特徵，我們先將語料庫依學習者不同母語分成六個子語料庫(sub-corpora)——英、日、韓、越、印尼、泰，從表 1 可以看到，母語為日語的語料量最多，佔全體語料的 24% 左右。而這六個母語背景的語料在 TOCFL 語料庫中排名前六名，約佔全體語料量的百分之七十左右，以下是每個子語料庫的總詞數以及佔全體語料的百分比統計訊息。

<sup>3</sup> 這裡是指舊版的 HSK 考試，當初設計雖有初、中、高等級，但僅高等考試需要寫作文。

<sup>4</sup> CEFR 的能力架構為三等六級，由初級到精通等級依序是 A1/A2, B1/B2, C1/C2；TOCFL 現階段語料庫中 A2 到 C1 等級語料的分布比例分別是 13%、48%、33%、6%。

表1. 六個子語料庫詞數及百分比訊息

	日語	英語	韓語	越南語	印尼語	泰語
詞數 (token)	271869	160400	110396	108999	87989	40071
佔全部語料 %	23.92	14.11	9.71	9.59	7.74	3.53

本研究應用語料庫檢索工具中常用的主題關鍵性 (keyword-keyness) 這個統計方法嘗試快速有效地找出特徵<sup>5</sup>。這個方法最早是應用在尋找文本中的主題詞，通過詞語的關鍵性分析 (keyness) 找出某一主題文本的詞語特徵 (William, 1976)。雖然早期多應用於文類或風格分析，但筆者假設利用主題關鍵性分析的統計概念—不尋常高頻 (unusually high) 或不尋常低頻 (unusually low) (Scott, 1997: 236)，找出不同子語料庫中相對高頻或低頻的關鍵性詞語 (Baker, 2006: 139)，應該可以呈現不同母語背景學習者的用詞特徵。

目前運用這關鍵性方法做的研究主要有兩種類型 (Rayson & Garside, 2000)，一是拿一個語料庫與一個較大 (標準) 語料庫的比較，目的是藉由顯著不同用法 (頻率) 來發現被觀察的語料庫的特徵；第二種是拿兩個差不多大小語料庫來比較，目的是想要找出可以區辨兩者 (語料庫) 的特徵。基本的原理都是利用統計的方法，比對兩個語料庫衍生出來的詞表，看看哪些詞是屬於不尋常高頻或不尋常低頻，進而從這些主題詞中來分析文本。不論哪個類型，原則上都要有一個參照語料庫 (reference corpus)，一個受觀察語料庫 (observed corpus)，參照語料庫要大於受觀察語料庫 (Scott & Tribble, 2006: 58)。因此第三節所呈現的結果，是使用兩種比對的結果，一是拿 114 萬詞學習者語料庫為參照語料庫，每個子語料庫為受觀察語料庫；一是分別拿日語背景語料庫、英語語料庫為參照語料庫<sup>6</sup>，其他子語料庫為受觀察語料庫。

分析的步驟是，先將不同母語 (英、日、韓、越、印尼、泰) 的語料庫製作成子語料庫 (受觀察語料庫)，進行斷詞工作後<sup>7</sup>，放入 AntConc (3.2.4w 版) 這個語料庫工具產出屬於各子語料庫的詞表，然後利用工具上 Keyword list 功能，計算出 Keyness，觀察這六個不同母語背景學習者的用詞特徵。圖 1 為進行計算日語學習者語料庫 keyness 的 AntConc 畫面<sup>8</sup>。

<sup>5</sup> 在語料庫工具 WordSmith (Scott, 1997, 2008) 或 AntConc (Anthony, 2009) 中，都可見 keyness 這個功能。

<sup>6</sup> 分別選定日語、英語為參照語料庫，主要是因為這兩者語料量是前二名，其他子語料庫的量則相對地小，不確定統計數據的可信度。

<sup>7</sup> 寫作文本斷詞是採用中央研究院的自動斷詞與詞性標註程式進行自動處理。參考網址：<http://ckipsvr.iis.sinica.edu.tw/>

<sup>8</sup> 畫面中第九個詞語「看電影」是自動斷詞的結果，其實語料庫中大多數都斷成「看」和「電影」兩個詞，這個斷詞結果不知為何原因和其他語料庫不同。雖然如此，並不影響我們後面的討論。

Concordance				Concordance Plot				File View				Clusters				Collocates				Word List				Keyword List										
Hits		Keyword Types Before Cut: 8469																Keyword Types After Cut: 1139																
Rank	Freq	Keyness	Keyword																															
1	611	284.239	日本																															
2	960	281.261	話																															
3	1984	186.159	所以																															
4	1440	102.539	時候																															
5	10846	90.364	我																															
6	104	87.758	日文																															
7	157	83.292	打工																															
8	16847	82.264	的																															
9	34	79.188	看電影																															
10	644	72.414	一起																															
11	380	67.654	公司																															
12	259	63.318	房間																															
13	113	60.400	日本人																															
14	351	60.321	吧																															
15	3247	52.192	你																															
16	269	50.391	房子																															
17	504	47.792	請																															
18	58	46.841	東京																															
19	44	45.131	愛文																															
20	101	44.698	理由																															
21	1247	41.524	那																															
22	990	41.076	可是																															
23	579	40.466	以後																															
24	401	40.161	然後																															

圖1. AntConc 3.2.4w 計算Keyness 畫面

### 3. 不同母語背景的用詞特徵

基本上，這六個子語料庫都是來自同一個大語料庫，寫作的主題都在一樣的範圍內，如果在不同子語料庫中可以看到不尋常高頻或不尋常低頻的現象，應該可以推測是受學習者母語這個因素的影響。

#### 3.1 關鍵性詞表結果

下表是依第 2 節研究方法中的陳述，為每個子語料庫做出關鍵詞表，然後依序列出前 20 高關鍵詞的資訊以便觀察<sup>9</sup>。從表 2 可以看到一個一致的現象，即這些不同母語背景的子語料庫顯現出來的高關鍵性詞都有代表該母語國的專有名詞，例如，日語語料庫中的「日本、日本人」，泰語語料庫中的「泰國、泰文、泰國人」等等。這個與整體語料庫的性質有關，因為考試作文題目中包括了介紹自己、說明喜好等與個人切身相關的主題，自然這些代表寫作者個人資料的訊息就會進入到寫作文本中。從這點，也可以初步確定這個關鍵性的方法，的確可以讓我們很快速輕易地看出屬於某個特定語料庫中的用詞特徵。

<sup>9</sup> 如圖 1 所示，第一個欄位顯示的是排名高低，可以找出前 20 名高關鍵詞。

表2. 六個子語料庫與整體學習者語料庫對比的前20 高關鍵性詞

子語料庫	前 20 高關鍵詞
日	日本 話 所以 時候 我 日文 打工 的 看電影 一起 公司 房間 日本人 吧 你 房子 請 東京 愛文 理由
英	我 會 她 你 個 美國 幫 感謝 台北 我們 相片兒 如果 王宜家 夢 李 吉他 見面 他們 可以 穿
韓	韓國 李廠長 不一樣 韓國人 時候 話 我 辣 的 錢包 泡菜 吧 以後 韓國菜 可是 中國菜 所以 還 寵物 容易
越	家長 你 您 小孩 工廠 都 電腦 攝影機 越南 影響到 上網 學習 生活 大家 處理 噪音 臭味 讓 每 很
印	印尼 學校 會 寵物 您 也 工廠 要 因為 在 們 或是 他 阿德 欣賞 可以 老師 畢業 狗 來
泰	泰國 他 她 就 書 今天 李孟 本 老師 平平 泰文 貴方 我 筆者 喬丹 去 幫 馬老師 泰國人 讓

下面表 3、表 4 的結果則是，我們分別將日語和英語做為參照語料庫，得出其他子語料庫與它們之間的高關鍵性詞表（一樣取前二十個）。由於這個研究方法的先決條件是參照語料庫要比受觀察語料庫大，所以下面表 3（以日語語料庫為參照）可以看到其他五個子語料的關鍵性詞，表 4 是以英語語料庫為參照，只能看到其他四個子語料庫的關鍵性，因為日語語料庫比英語大，無法作關鍵性分析。

表3. 與日語語料庫對比的前20 高關鍵性詞

子語料庫	前 20 高關鍵詞（類）
英	會 Nh 張愛文 和 美國 個 可以 Neu 他們 讓 Caa 我 一 就 幫 D 快 Nb VL 國家 她 [話、所以、吧、的]
韓	韓國 韓國人 李廠長 辣 錢包 泡菜 韓國菜 不一樣 媽 皮帶 膩 學生 D 主人 更 自然 種 熱 全 皮鞋 [個、然後、不過]
越	都 D 家長 越南 讓 COLONCATEGORY 就 發展 影響到 學習 可以 生活 Da 而 小孩 您 現象 很 所 電腦 [的、話、時候、我、所以]
印	印尼 D 會 在 人民 可以 此 要 能 他們 也 讓 或是 因為 Caa 學校 來 P 家長 牠們 [的、話、所以]
泰	泰國 就 他 讓 Nh Nb D 她 張愛文 VL 平平 李孟 泰文 書 幫 本 叫 喬丹 曼谷 影片 [的、話、時候、所以、吧]

表 4. 與英語語料庫對比的前 20 高關鍵性詞

子語料庫	前 20 高關鍵詞 (類)
韓	韓國 話 T 還 韓國人 寵物 吧 PAUSECATEGORY Ng 所以 李廠長 臺灣 你們 的 時候 辣 韓國菜 FW 台灣 以後 [個、會、代名詞(Nh)]
越	PAUSECATEGORY T 越南 您 家長 大家 學習 都 生活 處理 喔 電腦 還 而 這樣 Da 攝影機 影響到 網路 孩子 [的、代名詞(Nh)、個、會]
印	印尼 D 寵物 T 也 ETCCATEGORY 您 b 學校 你們 狗 畢業 來 們 喔 養 或是 阿德 事 出租 [代名詞(Nh)、個]
泰	泰國 T COMMACATEGORY 今天 他 就 你們 筆者 平平 李孟 泰文 月 泰國人 書 喬丹 曼谷 PAUSECATEGORY V 女生 有 [的、會、個]

在與日語語料庫和英語語料庫對比的資料表中，為了比較宏觀地來看彼此差異，我們在做關鍵性分析時，特別保留了詞類訊息（包括標點符號標記），因此在上面兩個表格中可以看到詞類訊息<sup>10</sup>。另外，也為了容易比對分析，將不尋常低頻的高關鍵性詞也一併選取前幾個具語法意義的詞彙用方括弧表示在表 3、表 4 中。例如，和日語學習者比較，英語學習者相對少用[所以、吧、...的話]<sup>11</sup>。

以下是我們歸納出不同母語背景學習者的用詞特徵。在歸納過程中，我們忽略表 3、4 中專有名詞（例如，「日本」）或具有實詞意義的詞（例如，「書」），而選擇具有語法或語用功能的詞語，以進一步分析並解釋可能的原因。

### 1. 母語為英語者的用詞特徵

- (1) 與整體對比，使用代名詞「我、她、你、我們、他們」的頻率相對突出。與日語對比，也可以看到代名詞 (Nh) 的高關鍵性。
- (2) 與整體對比，能願動詞「會、可以」的頻率相對突出。與日、韓、越、泰對比，「會」的使用頻率相對地高。
- (3) 與整體對比，使用量詞「個」的頻率相對突出。與日、韓、印尼、越、泰對比，也高。
- (4) 與整體對比，使用連接詞「如果」的頻率相對突出。與日語對比時，「所以」和「...的話」相對少用。
- (5) 與韓、越、印尼、泰對比，語助詞(T)相對少用。與日語對比，相對少用語助詞「吧」。

### 2. 母語為日語者的用詞特徵

- (1) 與整體對比，使用代名詞「我、你」的頻率相對突出。
- (2) 與整體對比，「的」的使用相對地突出。

<sup>10</sup> 這些詞類訊息標記代表的意義請參考詞庫小組（1995）。

<sup>11</sup> 表中方括弧內「的、話」是因為自動斷詞規則所致，筆者進語料庫實際觀察，發現「的、話」的高關鍵性主要是學習者使用「（如果）...的話」而來。

- (3) 與整體對比，連接詞語「所以、...的話」的使用頻率相對突出。與英、越、印尼、泰對比，也都高。
- (4) 與整體對比，語助詞「吧」的使用頻率相對地高。與英、泰對比，也都高。

### 3. 母語為韓語者的用詞特徵

- (1) 與整體對比，代名詞「我」的頻率相對突出。
- (2) 與整體對比，「的」的使用相對地突出。
- (3) 與整體對比，連接詞「所以、可是、...的話」的使用頻率相對突出。與英對比時，「所以」相對地高。
- (4) 與整體對比，語助詞「吧」的使用頻率相對地高。與英對比時，也高。
- (5) 與整體對比，副詞「還」的使用頻率相對突出。與日語對比時，副詞「更」相對地高；與英語對比時，副詞「還」相對地高。

### 4. 母語為越語者的用詞特徵

- (1) 與整體對比，代名詞「你、您」的頻率相對突出。
- (2) 與日語對比時，能願動詞「可以」的使用頻率相對突出。
- (3) 與英語對比時，語助詞「喔」的使用頻率相對地高。
- (4) 與整體對比，副詞「都」的使用頻率相對突出。與日語對比時，副詞使用頻率相對地高；與英語對比時，副詞「都、還」相對地高。

### 5. 母語為印尼語者的用詞特徵

- (1) 與整體對比，代名詞「您、他」的頻率相對突出。
- (2) 與整體對比，能願動詞「會、可以」的頻率相對突出。與日語對比時，「會、可以、要、能」相對地高。
- (3) 與整體對比，使用連接詞「或是」的頻率相對突出。與日語對比時，「或是、因為」相對地多。與英語對比時，「或是」相對地高。
- (4) 與整體對比，副詞「也」相對突出。與英、日對比時，也高。
- (5) 與整體對比，後綴詞「們」相對突出。與英語對比，也高。
- (6) 與英語對比時，語助詞「喔」相對地高。

### 6. 母語為泰語者的用詞特徵

- (1) 與整體對比，代名詞「他、她、我、筆者」的頻率相對突出。與日語對比時，代名詞使用頻率相對地高。與英語對比時，「他、你們、筆者」相對地高<sup>12</sup>。
- (2) 與整體對比，量詞「本」的頻率相對突出。與日語對比時，也高。

<sup>12</sup> 我們實際進泰語子語料庫查詢，使用「筆者」都是同一位學習者（B1 等級），屬個別現象。

- (3) 與整體對比，副詞「就」相對突出。與英、日對比時，也高。
- (4) 與英語對比時，語助詞（T）的使用頻率突出。

## 3.2 分析討論

根據上述的歸納，為了較有系統地來討論不同母語背景學習者用詞的特徵與可能原因，下面以詞類來分項敘述。

### 3.2.1 代名詞

從關鍵性分析結果來看，英語代名詞(表中的 Nh) 的使用相對於其他五個語言背景學習者特別顯著，除了高關鍵性的代名詞種類最多，與日語相比，英語學習者的代名詞這個詞類的關鍵性也是突出。韓、越、印尼學習者則沒有這個特色。

代名詞在篇章中主要起銜接作用。不少研究者或語言教師都會指出初階學習者的特徵之一就是不會使用零代詞( zero anaphora)，使得篇章結構鬆散、連貫性差( 尚彞強, 2001: 54)，而多數實證研究的對象都是歐美或西方學習者，例如，Jin (1994) 從語言類型的角度，探討母語為英語的學習者，在學習屬於主題顯著的語言（華語），有將英語的主語結構特徵遷移到第二語言的現象；周曉芳（2011）則是針對歐美學生敘述語篇中的回指習得研究指出學習者呈現代詞多用、零代詞嚴重少用的情況。研究者對歐美學習者代詞習得感興趣的原因，主要應該是從英漢語言類型差異的角度出發。研究者會從主題顯著語言/主語顯著語言（Li & Thompson, 1976）、代詞脫落（pro-drop）語言/非代詞脫落語言來解釋學習者過渡語言的現象。

我們從語料分析數據中得出整體學習者有代名詞多用的情況，但如果以日語做為參照語料庫，則只有英語和泰語的代名詞類出現關鍵性數值。表示英語與泰語學習者代名詞多用的情況是比其他四個語言背景的學習者來得顯著。由於我們所觀察的這六個語言背景中，只有英語屬於非代詞脫落/主語顯著語言，其他五個亞洲語言都屬於代詞脫落/主題顯著語言（Chomsky, 1981: 284, footnote 47; Li and Thompson, 1976）。這裡泰語母語者在代名詞部分的突出表現則顯得與本研究中其他亞洲語言的表現不一致，我們嘗試從前人的研究中找線索，但只有關於泰語學習者在回指偏誤這方面的統計結果，例如，楊帆（2011: 14）、林雪鳳（2008: 62）指出指稱偏誤中，零形式的偏誤情形最多，也就是學習者不善於使用零形式，而使得篇章行文不流暢；並沒有從語言類型或對比的角度的分析偏誤的原因，前人分析這方面偏誤的原因，都偏重在教學上，普遍認為教材或教學沒有著重在這方面做有系統的安排。

不過，泰語的代名詞數量相當驚人，會因說話者與聽話者的性別和相對關係等有所不同，光是「我、你」的說法就各有十餘種<sup>13</sup>。或許泰國學習者對於代名詞的使用特別

<sup>13</sup> 引自網路上文件：[thai.world68.com/upfiles/泰語人稱代詞詳解.doc](http://thai.world68.com/upfiles/泰語人稱代詞詳解.doc)

有意識，也就是說，母語者的詞類系統對於二語習得也是一個影響因素<sup>14</sup>，當然這需要未來更多的研究來驗證。

另外我們也發現無論與整體語料庫或日語、英語語料對比時，越南母語者代名詞「您」都具關鍵性。前人的文獻顯示，越南語的人稱代詞系統很複雜，而且多數帶有情感色彩，例如不同形式可以表示親密、尊敬、中性、輕蔑等態度（阮氏懷芳, 2008: 46），漢語則大不相同，除了第二人稱「您」帶有尊敬色彩，其他人稱都是不帶感情色彩的中性詞。推測越南學習者在使用人稱代詞時，特別有意識地，會根據場合、身份，選用第二人稱「您」。我們實際進入越南語子語料庫驗證了我們的推測，凡寫作主題中牽涉到給老師、父母、工廠廠長、政府公務人員等便條或信件的內容，需使用第二人稱代詞時，越南學習者幾乎都選用敬語形式，包括「您、您倆、您們」。

### 3.2.2 連接詞

在連接詞的部分，本研究先就「因為/所以、如果/...的話」這兩組呈現高關鍵性數值進一步分析。

#### 3.2.2.1 「因為/所以」

張莉萍（2013a）觀察本國人與學習者連接詞使用分佈時，即注意到學習者使用「所以」的頻率較「因為」高，與本國人不同。現在關鍵性的分析呈現日語和韓語者使用「所以」的高關鍵性。由於語料庫中絕大多數初階學習者是在台灣接受華語課程，我們可以排除教學、教材這些變數對學習者的影響，加上在華語教學中，表因果關係的「因為、所以」都是同時成對授予學生的，因此，我們推測日語、韓語學習者的這個使用傾向受母語影響的可能性不低。

下表是針對「因為、所以」這兩個連接詞，英、日、韓這三個母語背景學習者各等級的使用狀況。表 5 顯示，三類學習者都有一個共同趨勢，即這兩個詞的使用率隨著能力的提高而降低。如果參照本國人的使用率<sup>15</sup>，「所以」是 0.13-0.56，「因為」是 0.16-0.85，可以看出學習者使用「因為」的頻率和本國人較接近，「所以」的使用率則要到 B2（以上）能力比較接近本國人。

<sup>14</sup> Jarvis, Gastaneda-Jimenez and Nielsen (2012:61) 研究中指出不少學習者母語影響外語學習的例證，例如，芬蘭語學習者明顯較其他母語者少用英語的 *a, the*，是因為芬蘭語中沒有冠詞(articles) 這一個語法系統。另一例是，芬蘭語中沒有區分性別的代名詞系統，從語料庫數據中也可以看到以芬蘭語為母語的學習者在英語 *he, she* 的使用上，相對較其他母語者來得少。

<sup>15</sup> 為免頻率訊息受書面或正式語體的因素影響，這裡的使用頻率是同時引自中研院平衡語料庫詞頻頻率（<http://elearning.ling.sinica.edu.tw/CWordfreq.html>）以及對話語料庫頻率訊息（<http://mmc.sinica.edu.tw>）。對話語料庫是由 30 個自由對話，26 個地圖導引對話與 29 個主題對話的內容整理而成，約計 40 萬詞，該語料庫詞表於 2012 年 4 月公開。平衡語料庫則是台灣第一個對外開放的大型語料庫，以書面語料居多，1995 年開放 1.0 版語料庫是 200 萬詞，本研究使用的是 3.0 版 500 萬詞（<http://app.sinica.edu.tw/kiwi/mkiwi/>）。

表5. 「所以、因為」英、日、韓語學習者使用情況

		A2 (%)	B1 (%)	B2 (%)	C1 (%)
所以	日 だから	1.4609 <sup>16</sup>	1.0139	0.5603	0.4166
	韓	1.4859	1.0445	0.3855	0.1543
	英	0.9783	0.5923	0.2912	0.2419
因為	日 ので	0.6598	0.5564	0.4854	0.2406
	韓	0.7676	0.6017	0.3905	0.3305
	英	0.8145	0.6282	0.5670	0.3892

這其中又以日語學習者使用「所以」的頻率最為突出。英語學習者使用「因為、所以」這組連接詞的傾向和目標語人士較接近，即「因為」的頻率高於「所以」；韓語學習者則到 B2 等級才漸漸有這個傾向(0.3905:0.3855)；日語學習者則一直呈現「所以」高於「因為」的結果。

周剛(2001)在〈漢、英、日語連詞語序對比研究及其語言類型學意義〉一文中舉了因果目的關係連詞的例子如下<sup>17</sup>：

(1) 因為他的汽車在路上出了事，所以來晚了。

(1') He came late because his car broke down on the way.

(1'')a. 彼 は 自動車 が 途中 事故 に あったので、来る の が 遅れた。

Kare wa zidousha ga tochu jiko ni attanode, kuru no ga okureta

b. 彼 は 自動車 が 途中 事故 に あった。だから 来る の が 遅れた。

Kare wa zidousha ga tochu jiko ni atta dakara kuru no ga okureta

(引自周剛, 2001: 46 例句 20)

說明日語表因果目的的連詞可以用接續助詞，如(1'')a 的ので（因為），也可以用接續詞，如(1'')b 的だから（所以），接續助詞都是後置先行的<sup>18</sup>；接續詞都是前置後續的。

<sup>16</sup> 數據的算法是將日語者使用「所以」詞數除以日語者總詞數(e.g. 403/271869)，表示母語為日語學習者，使用「所以」的比例。

<sup>17</sup> 引用之原文例子沒有附逐字翻譯之羅馬拼音，為利閱讀，我們補充於此。後面例句(2)同此處理方式。

<sup>18</sup> 該文所探討的語序敘述如右：「連詞的語序應該包括兩種，一種是指連詞處於所連接的成分的前後位置，我們稱之為前置和後置；另一種是指連詞處於話語中的語序，有固定和不固定之分，我們稱之為定序和不定序，定序連詞中又可以分為先行和後續兩種。先行連詞中有些詞由於語用因素，可與所連接的成分移動至後續小句的位置，有些則不能，我們把它們稱之為可後移與

而且表示因果關係的連詞，只能用單個詞，不能用兩個詞搭配的關聯形式（像「因為...所以...」）。

而漢、日表示因果關係的語序，都是原因在前，結果在後，如果要強調結果或是做為事後補充語，有時表示結果的子句，也會出現在前（Chao, 1968: 132）<sup>19</sup>。我們進一步查詢ので（因為）和だから（所以）這兩個詞在日語的頻率訊息，發現だから的頻率遠高於ので。所以我們推測日語學習者使用「所以」頻率較高的原因有兩個，一是因為日語在這類連接句中，只能選擇單一連接詞語，而だから（所以）的頻率又高於ので（因為）許多，這個習慣遷移到漢語中<sup>20</sup>，使用「所以」的頻率相較於其他不同母語背景學習者，才會那麼突出；二是語言類型的影響，我們假設學習者傾向選擇和目標語一致的前置連詞來使用，だから位於（小）句首位置，同漢語連詞的典型位置相同。不論是哪一個因素，這個現象一直持續，在高階學習者身上仍可見。韓語表達因果關係的方法與形式與日語大同小異，大多也都使用一個連詞，不用成對的形式，呈現在漢語學習上，用詞特徵與日語學習者大致相同。

### 3.2.2.2 「如果...的話」

漢語表示假設條件關係的連接詞語，最常見的就是「如果...（的話）」。其中「如果」的頻率不論在平衡語料庫或對話語料庫中都高於「的話」，「...的話」並不是漢語典型的連接形式<sup>21</sup>，漢語典型的連接詞語多放在句首位置，不論是主要句子或從屬句子。但關鍵性分析中，日語、韓語都呈現「的話」的高關鍵性。於是我們進一步篩選統計了英、日、韓三種語言單用「如果」、單用「的話」以及使用「如果...的話」這三種不同形式，結果如下表：

表6. 英、日、韓學習者「如果...的話」使用比例統計

	如果...的話	如果...	...的話	小計
英	15.51%	71.62%	12.87%	100%
日	45.45%	18.18%	36.36%	100%
韓	29.60%	25.37%	45.03%	100%
本國人	6.51%	80.72%	12.77%	100%

不可後移連詞。反之，後續連詞中有些詞由於語用因素，可與所連接的成分移動至先行小句的位置，有些則不能，我們把它們稱之為可前移與不可前移連詞。」（引自周剛, 2001: 42）

<sup>19</sup> Tai (1985) 解釋「先說因，後說果」是漢語的自然語序(natural order)；「先說果，後說因」則是凸顯語序(salient order)。

<sup>20</sup> 關於 L1 頻率的遷移，在 Jarvis and Crossley (2012) 主編的書中，也有不少討論，例如，瑞典語與芬蘭語學習者在使用英語 come 這個詞彙的頻率高於葡萄牙語和西班牙語學習者(p.62)，是受到他們母語中相對應詞彙頻率的影響。葡語和西語母語中使用 llegar (arrive) 較高，反映在使用英語時，arrive 的使用率高於 come。

<sup>21</sup> 「的話」形式出現於元朝（余志鴻, 1992; 祖生利, 2002），是受阿爾泰語言影響（感謝曹逢甫教授提供此資訊）。

從上表可以發現，對於這三種形式的使用，英語母語者選擇單用「如果」這個連接詞的比例最高（71.62%），也最接近本國人使用情況（80.72%）<sup>22</sup>；韓語者選擇單用「的話」比例最高，日語者選擇「如果...的話」比例最高。而日韓共同點是單用「如果」的情況最少，選擇含有「的話」的表達形式佔了75%以上，日語更高達82%。我們從前人文獻知道，日語中一般表達假設條件關係的連詞只用接續助詞，如下面(2'')例句中的“たら”。

(2) 如果她拒絕了，那麼你怎麼辦？

(2') In case she refuses, what will you do?

(2'') もし 彼女 が 断つたら、 あなた は どうする の ですか。

Moshi kanojo ga kotowaruttara anata wa dosuruno desu ka.

（引自周剛, 2001: 49 例句 36）

推測日語母語者受日語條件關係表達中，都用接續助詞（後置黏著詞）的影響，傾向使用漢語中性質接近的「的話」來表達假設語義。韓語在連接詞語的表現和日語相近，大都使用詞尾成分，接在詞幹或時間詞尾後<sup>23</sup>。因此日韓學習者的表現較一致，與英語學習者選用連接成分（形式）差異較大。

筆者推測日韓學習者傾向於使用後置成份，因此另外觀察「除了...以外/之外/外」這句式，統計數據顯示日韓學習者保留使用後置成份「以外/之外/外」的比例遠高於英語背景者；日語有93.65%、韓語有89.23%、英語有69.23%，雖然保留後置成份的比例都很高，不過，英語背景者不用後置成份的情形還是明顯高於日韓學習者。

對表達同一個語義的不同形式，日韓學習者似乎有志一同的選用和自己母語相似的表達形式，而非選用目標語人士較高頻使用的形式。從這兩組連接詞的表現，似乎也可以看出日韓語言類型接近，呈現在學習漢語的表現形式也接近<sup>24</sup>。

<sup>22</sup> 這個統計結果是根據中央研究院漢語平衡語料庫 4.0 版 (<http://140.109.19.114/>) 檢索得來。

<sup>23</sup> 郭麗娟 (2011: 26) 指出漢語假設連詞「的話」和韓語「으면 (的話)」都放在動詞的後面，位置相同。

<sup>24</sup> 從構詞類型的角度來看 (Lehmann, 1978: 219)，日語、韓語偏向黏著型語言 (agglutinative language)，是透過詞綴加入詞幹以增加意義或改變語法功能的語言。英語偏向屈折型語言 (inflectional language)，透過詞形變化表示意義和語法功能的改變。漢語偏向孤立型語言 (isolating language)，透過語序和功能詞表達語法關係。

### 3.2.3 能願動詞

在能願動詞的部分，英語母語者的「會、可以」具高關鍵性；印尼母語者的「會、要、可以」具高關鍵性。如果以日語為參照語料庫，則印尼和越南學習者對能願動詞的使用頻率相對突出。能願動詞這一類無論在語言學或華語教學中，都有相當多的討論，在華語學習上，也是一個公認的難點（鄧守信, 2009: 113）。一般教師或研究者多從能願動詞混用的角度來分析原因（賴鵬, 2006），葉信鴻（2009）則從 HSK 動態作文語料庫中，實際統計「會、能、可以、想、要」等五個主要能願動詞的偏誤類型發現，缺漏 (omission) 與誤加 (addition) 這兩類遠較錯用 (mis-selection) 與錯序 (mis-ordering) 等類型比例來得高，一般文獻中所關心的混用或誤用情況反而不是偏誤較高的類型。

由於筆者不熟悉印尼語，也沒找到相關探討印尼學習者學習漢語能願動詞的研究，因此在這裡，僅能就英語學習者來討論。從語言類型的角度來解釋，似乎可以推測因為英語助動詞與漢語能願動詞類有相近的語義和語法範疇，幾乎也都有可以對應的詞語，雖然不一定是一對一，可能是一對多，但存在著相應的類，對英語學習者來說應該比母語沒有獨立的這類學習者較不會有缺漏或誤加的現象。我們從英語語料庫「會、可以」相較於其他子語料庫的高關鍵性，可以大膽推測其他語種學習者缺漏情況較英語學習者顯著。由於缺漏，需要分析者一筆一筆觀察標記，無法由機器判讀，因此限於時間、人力，在這個研究裡，僅觀察 B2 等級（中高級，相當於 HSK 語料等級）子語料庫中「會」的使用情況，結果如表 7 所示：

表 7. TOCFL 語料庫 B2 等級子語料庫「會」的使用情況

	篇數	使用次數	偏誤小計	缺漏次數 %		誤加次數 %		錯用次數 %	
日	260	396 (0.53)	305	204	66.89	81	26.56	20	6.56
英	122	418 (1.07)	127	54	42.52	37	29.13	36	28.35
韓	130	268 (0.68)	42	16	38.10	9	21.43	17	40.48
泰	41	96 (0.89)	28	10	35.71	4	14.29	14	50.00
印尼	112	335 (1.05)	38	2	5.26	3	7.89	33	86.84
越	188	401 (0.73)	28	5	15.15	12	36.36	16	48.48

日語學習者缺漏「會」的比例最高，英語學習者次之，並不如筆者先前推測，因此英語這個高關鍵性的表現，未來還需要進一步探討。

從表 7 也可以看出，不同母語者「會」的偏誤類型表現不一，但值得我們注意的是，除了傳統教學者或研究者關注的混用（錯用）情況，學習者在該用卻不用的情形也相當顯著，而且這情況是存在已趨近本國人表現的 B2 學習者身上，可見「會」或其他能願動詞的使用問題，包括該用不用或誤加等等，也值得研究者或教學者深入發掘探討。以下僅舉幾例語料庫中缺漏偏誤供參（方括弧內是筆者加入的）。

- (3) \*希望李先生能夠了解一個小小的消費者的情境，並相信政府當局 [會] 妥善處理。  
(B2 程度學習者，以下 2 例皆是)
- (4) \*如果製造「有錢就可以解決問題」這樣的世界，人的心 [會] 越來越空虛。
- (5) \*而且長期電腦使用 [會] 對身體帶來很多不好的影響，例如視力不良，肩膀酸痛等。

### 3.2.4 語氣詞

這裡的語氣詞指的是放在句尾的語氣助詞，如，「啊、吧、嘛」等等。在高關鍵性的分析結果中，可以清楚看出英語和其他五個亞洲語言在這方面明顯的差異。也就是，英語學習者跟整體學習者比較下，相對少用語助詞。日、韓學習者則是相對多用語助詞「吧」；越南語、印尼語、泰語則是和英語對比時，才呈現出語助詞的關鍵性。從類型學的觀點來看，英語沒有句尾語氣助詞這類詞語，多用助動詞或疑問句型來表達這樣的情態；而日、韓等東亞語言幾乎都有表達情態的助詞，推測這是英語背景學習者少用的主要原因。徐晶凝（2003）從日漢對譯的語料中來分析「吧」的語法分佈情形，指出漢語「吧」基本上的出發點是在於說話者對命題真值的關心，可以從日語對譯的某種形式「ね 疑問句」得出，這是一種預料聽話者對詢問內容能給以肯定性確認的問句；然而日語說話者在交際時，頻繁使用語氣助詞，例如“ね”，並不是為了讓對方針對他提供的命題信息進行確認，而是一種潤滑雙方關係的手段。這個觀點，我們從日語學習者語料中也得到證實，日語母語者的語言習慣讓「吧」的使用頻率較其他母語背景者高，而且多用在表示禮貌的委婉語氣，而非想得到訊息的確認。例如，下面這段擷取自日籍學習者的考試作文語料，可以看出使不使用「吧」並沒有合法性的問題，使用了語氣助詞具有緩和語氣、促進雙方情感交流的效果。

我已經詳細地看過了妳的信。關於妳同學的錢包不見了這事，我可以理解妳的心情，同時我也替你感到憤怒。無論你的位子在不在他的旁邊，總不該把同學當做小偷來對待**吧**！...後來，他知道你背包裡沒有他的錢包，他向妳道了歉嗎？...我不知道平常妳跟他好不好，但我想，他是一時焦急才這麼做的**吧**。下次碰到他時，妳可以問一下他的錢包找到了沒有。並且可以告訴他，妳對他的對待多麼的傷心。...聽到妳講的話，他有什麼道歉，那妳就原諒他**吧**。...妳剩下的留學期間也不多了**吧**，好好學習，好好玩啊！

（資料來源：TOCFL 語料庫日籍 B2 學習者）

韓語的語氣表現方式與日語差不多，主要是以終結詞尾來表達說話者的態度和意圖（謝瑩 2011: 26）<sup>25</sup>，與漢語的句末語氣相似。謝瑩（2011: 27）認為韓國學生出於對於禮貌策略的敏感性，會主動選擇在句末添加語氣詞以達到表示禮貌的目的，於是出現語氣詞多用的偏誤，其中包括了「吧」的使用。本研究尚未考察語氣詞偏誤情況，不過，從上述日語學習者的這段語料，可以知道如果多用不至於構成偏誤，可以視為日語學習者，甚至是東亞學習者的語體風格。至於多用的因素，那可能是文化或語用因素導致學習者遷移而多用句尾語氣詞。至少從關鍵性分析結果來看，東亞語系背景學生在語氣詞使用的表現上，是較英語背景學生突出許多。

### 3.2.5 量詞

在量詞部分，英語學習者「個」的高關鍵性也是相當突出的<sup>26</sup>。相較於其他五個亞洲語言，英語雖也有度量衡詞或以容器表達的數量詞語，但對於歐美學生來說，漢語的量詞在英語並沒有相對應的形式，如，一「本」書、一「匹」馬。那麼，為什麼英語學習者反而在量詞「個」呈現不尋常高頻的現象，我們推測是教學時，這個語法上的結構「數詞+量詞+名詞」並不難，學習者知道要使用量詞，只是不一定能掌握量詞和名詞的搭配，量詞不一定用得正確，而「個」是漢語中最通用的量詞，如果不確定使用哪個量詞或學習者能力還不及，多會選擇「個」，因此歐美學生使用「個」的頻率可能還高於母語人士，這一點我們從張博等人（2008: 91）從中介語語料庫中得到的調查結果也可印證，他們從 1636 個歐美學生使用量詞的用例中，得出「個」的用例有 1076 句，佔總數的 66%，其中在名量搭配這類偏誤中，「個」的泛化，佔了 52.53%，也是比例最高的。

另外，我們從探討日、韓、泰學習者學習漢語量詞的文獻中發現，雖然這些語言有量詞這一類，但是名量的搭配、量詞的位置或使用量詞的條件與漢語不盡相同（宮下研介, 2011；金龍勛, 2011；宋帆, 2008），對這些學習者而言，難度並不比歐美學習者來得低，我們從張博等人（2008: 79-97）統計韓國學生和歐美學生的正誤用例，前者偏誤的比例（11%）還比後者（7.3%）來得高，也可以看出量詞不只對歐美學習者而言，是難點；對整體學習者而言，也是一樣的。

表 8 是 TOCFL 語料庫中六個子語料庫量詞的使用情況，可以看出英語學習者使用「個」這個量詞佔所有量詞用例的 46% 左右，較其他母語學習者高了三至九個百分比左右。這個結果與張博等人利用英、韓中介語語料庫所調查出來的結果一致，不過，他們並沒有觀察到英語學習者較其他語言背景學習者在使用「個」的頻率突出表現。

<sup>25</sup> 終結詞尾是指位於一句話終了時，處於最後一個謂語的詞尾。除了表示這句話已經結束，還表達說話者的態度與意圖。

<sup>26</sup> 泰國學習者「本」也具有高關鍵性，不過，從表 2 中，可以知道是受「書」高頻率的影響，屬個別情況，這裡不做討論。

表8. 不同母語背景學習者量詞「個」使用情況

	英	日	韓	越	泰	印尼
使用量詞種類	112	135	84	97	64	84
「個」使用次數	2405	3164	1099	1289	486	1086
使用量詞次數	5256	8162	2973	3410	1342	2541
「個」所佔比例	45.76%	38.77%	36.97%	37.80%	36.21%	42.74%

究竟是泛化（誤用）或是誤加？抑是英語學習者較日、韓語學習者更有意識地使用量詞？則需要進一步觀察分析。初步我們利用偏誤檢索工具觀察「個」誤加的情況<sup>27</sup>，發現在 33 個誤加用例中，西方學生用例佔了 60%；東方學生用例佔了 40%。似乎西方學生誤加的情況較東方學生嚴重，不過，這個誤加的情況不只是量詞「個」的問題，而是英語學習者會在不需要加數量結構時，使用「一個」，推測與英語冠詞“a/the”的表達形式有關，所以會造出「我看見一個漂亮的風景」這樣的句子。

#### 4. 結語與研究限制

本研究利用主題關鍵性方法觀察不同母語背景學習者在選詞、用詞上的特徵。這個方法讓我們快速、有目標地進一步分析不同母語者的用詞特徵，這些特徵包括英語學習者代名詞和能願動詞「會、可以」的使用突出性、相對少用句尾語氣助詞、多用量詞「個」；日韓學習者在假設句使用上優先選用「的話」形式、因果句優先選用「所以」等等。

這些關鍵性特徵大致上都可以從語言類型、語言結構、詞類、語用（文化）等面向上給予解釋，包括：（1）在選詞上，可以看出學習者傾向於選擇與自己母語類型接近的語言形式，例如，日語者在表達漢語因果關係時，優先選擇了「所以」，因為日語裡沒有像漢語中「因為...所以...」這樣成對的形式，只能用單一詞語，而日語中的だから，非黏著性而且位於句首的位置，最接近漢語的「所以」；相對於「因為」のので，則是黏著語素，在日語的使用頻率上也遠低於だから。另外，還有假設條件句中「的話」形式的高關鍵性。（2）母語中是否有與漢語相對應的詞類，似乎也是影響選詞、用詞的因素之一，例如英語背景學習者有助動詞這類，以及沒有句尾語助詞這類所呈現的關鍵性表現。（3）母語中的文化因素，也會遷移到漢語的使用，例如，越南學習者使用漢語第二人稱敬語形式的高關鍵性。這些資訊都可以經過轉化做為提升教學成效的基礎。

這個研究原來只是個嘗試，幸運的是看到應用主題關鍵性方法的可行性，可以有效地找出不同母語背景學習者的語言特徵，也可以透過這個方法，讓二語習得研究者有針對性地探討分析。除此之外，主題關鍵性分析方法未來也可以嘗試應用在機器自動辨識文本寫作者的國籍訊息上（Jarvis & Crossley, 2012），進一步偵測出可能的偏誤（尤其是學習者誤加或遺漏成分）。

<sup>27</sup> 檢索網址為 <http://kitty.2y.idv.tw/~hjchen/cwrite-error/>

當然這個研究由於現階段語料量不大（一百七十多萬字），因此無法將不同母語者的語料再依不同能力（等級）區分來觀察，只能等待語料庫日漸茁壯後，可作更精緻的分析，例如，觀察學習者這些顯著特徵是否隨著能力不同而產生變化。最後，本研究限於時間與人力，僅整理出前 20 高關鍵性詞語，實際上可能還有更多待發現的關鍵特徵。即使是已經整理出的關鍵特徵，部分也因為限於筆者自身對這些外語的知識不足或找不到相關對比研究文獻而無法進一步解釋，或做深入的分析探討，例如，能願動詞「可以」和一些高頻副詞，未來需要更多研究者投入開發。

### 參考文獻

- Anthony, L. (2009). Issues in the design and development of software tools for corpus studies: The case for collaboration. *Contemporary corpus linguistics*, ed. by P. Baker, 87-104. London, UK: Continuum Press.
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Chao, Y.-R. (1968). *A grammar of spoken Chinese*. Berkeley/Los Angeles: University of California Press.
- Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Holland: Foris Publications. Reprint. 7th Edition. Berlin and New York: Mouton de Gruyter, 1993.
- Council of Europe [COE]. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Díaz-Negrillo, A., & Fernández-Domínguez, J. (2006). Error tagging systems for learner corpora. *Resla*, 19, 83-102.
- Granger, S. (Ed.). (1998). *Learner English on computer*. London & New York: Longman.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition and foreign teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hawkins, J. A., & Buttery, P. (2009). Using learner language from corpora to profile levels of proficiency: Insights from the English profile programme. *Language testing matters: Investigating the wider social and educational impact of assessment, Studies in Language Testing*, ed. by L. Taylor, and C. J. Weir, 31, 158-175. Cambridge: UCLES/Cambridge University Press.
- Jarvis, S., & Crossley, S. A. (Eds.) (2012). Approaching language transfer through text classification: Explorations in the detection-based approach. Bristol, UK: Multilingual Matters.
- Jarvis, S., Gastaneda-Jimenez, G., & Nielsen, R.. (2012). Detecting L2 Writers' L1s on the Basis of Their Lexical Styles. *Approaching language transfer through text classification: Explorations in the detection-based approach*, eds. by Jarvis and Crossley, 34-70. Bristol, UK: Multilingual Matters.
- Jin, H. G. (1994). Topic-prominence and subject-prominence in L2 acquisition: Evidence of English-to-Chinese typological transfer. *Language learning*, 44, 101-122.

- Lehmann, W. P. (Ed.). (1978). *Syntactic Typology: Studies in the phenomenology of language*. Hassocks: Harvester Press.
- Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. *Subject and topic*, ed. by Charles N. Li, 457-489. New York: Academic Press.
- Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *proceedings of the workshop on comparing corpora*, held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000), 1-6. Hongkong.
- Scott, M. (1997). PC analysis of key words - and key key words. *System*, 25(2), 233-245.
- Scott, M. (2008). *WordSmith tools. Version 5*. Liverpool: Lexical Analysis Software.
- Scott, M., & Tribble, C. (2006). *Textual Patterns: Keywords and corpus analysis in language education* [Studies in Corpus Linguistics 22]. Amsterdam: John Benjamins.
- Williams, R. (1976). *Keywords: a vocabulary of culture and society*. New York: Oxford University Press.
- 余志鴻 (1992)。元代漢語的後置詞系統。《民族語文》，3，1-10。
- 宋帆 (2008)。《漢泰語量詞比較研究和泰語量詞教學》。上海市：上海外國語大學語言學及應用語言學專業碩士論文（未出版）。
- 肖奚強 (2001)。外國學生照應偏誤分析。《漢語學習》，2001 年第 1 期，50-54。
- 阮氏懷芳 (2008)。《漢越人稱代詞對比研究》。桂林：廣西師範大學語言學及應用語言學專業碩士論文（未出版）。
- 林雪鳳 (2008)。《泰國初級漢語學習者敘述體語篇銜接之研究》。廈門：廈門大學語言學及應用語言學專業碩士論文（未出版）。
- 周剛 (2001)。漢、英、日語連詞語序對比研究及其語言類型學意義。《語言教學與研究》，2001 年第 5 期，42-54。
- 周曉芳 (2011)。歐美學生敘述語篇中的回指習得研究過程。《世界漢語教學》，25(3)，422-432。
- 宮下研介 (2011)。《漢日量詞對比研究—對日漢語中的量詞教學》。黑龍江：黑龍江大學漢語國際教育專業碩士論文（未出版）。
- 金龍助 (2011)。《中韓量詞對比研究》。蘇州：蘇州大學漢語國際教育專業碩士論文（未出版）。
- 徐晶凝 (2003)。語氣助詞“吧”的情態解釋。《北京大學學報（哲學社會科學版）》，40(4)，143-148。
- 祖生利 (2002)。元代白話碑文中助詞的特殊用法。《中國語文》，5，459-480。
- 張莉萍 (2012)。對應於歐洲共同架構的華語詞彙量。《華語文教學研究》，9(2)，77-96。
- 張莉萍 (2013a)。TOCFL 作文語料庫的建置與應用，載於崔希亮、張寶林（主編），《第二屆漢語中介語語料庫建設與應用國際學術討論會論文選集》（頁 141-152）。北京：北京語言大學出版社。
- 張莉萍 (2013b)。華語學習者句式使用情況分析。發表於 2013 台灣華語文教學學會年會暨國際學術研討會（2013.12），高雄：文藻外語大學。

- 張博等人（2008）。*基於中介語料庫的漢語詞彙與專題研究*。北京市：北京大學出版社。
- 詞庫小組（1995）。中央研究院漢語語料庫的內容與說明。*技術報告第95-02/98-04號*，中央研究院。
- 葉信鴻（2009）。*現代漢語助動詞的界定與教學應用*。台北：國立台灣師範大學華語文教學研究所碩士論文（未出版）。
- 郭麗娟（2011）。*中高級水平韓國留學生漢語連詞使用情況研究*。南京：南京師範大學漢語國際文化教育學院碩士論文（未出版）。
- 楊帆（2011）。*泰國學生漢語語篇銜接手段偏誤分析及教學研究*。重慶市：西南大學語言學及應用語言學專業碩士論文（未出版）。
- 鄧守信（2009）。*對外漢語教學語法（修訂二版）*。臺北市：文鶴出版社。
- 賴鵬（2006）。漢語能願動詞語際遷移偏誤生成原因初探。*語言教學與研究*，2006(5)，67-74。
- 謝瑩（2011）。*高級水平韓國留學生漢語語氣詞“吧、嗎、呢”偏誤分析*。上海：華東師範大學漢語國際教育碩士專業碩士論文（未出版）。