

評 *Normalization in Translation: Corpus-based  
Diachronic Research into Twentieth-century  
English-Chinese Fictional Translation*  
兼論語料庫翻譯學新的研究趨勢

高照明

書名：*Normalization in Translation: Corpus-based Diachronic  
Research into Twentieth-century English-Chinese Fictional Translation*

書籍作者：Yun Xia

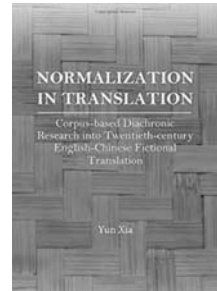
出版社：Cambridge Scholars Publishing

出版日期：2014年8月1日

ISBN：1443860379

總頁數：190頁

售價：51.87美元



傳統的翻譯學研究主要探討原文和譯文之間的對應，歐美翻譯學者 90 年代開始另闢蹊徑，將研究的焦點從原文和譯文之間的對等關係（equivalence）轉移到譯文的普遍性特徵（又譯為翻譯共性）（translation universals）。首開先河的是莫娜貝克（Mona Baker），Baker（1993）建立了語料庫翻譯學的方法論，利用語料庫語言學的分析方法，透過大量的原文語料庫和譯文語料庫來探索翻譯的共性。Baker 歸納出譯文通常有以下四個共同的特徵：簡化（simplification）、顯化（explicitation）、範化（normalization）、整齊化（leveling out）〔本文對這些術語的翻譯

參考王克非與黃立波（2008）〕。簡化是指譯文的詞彙及結構通常較原文簡單，顯化是指譯文語義跟原文相比通常更清楚，而範化則是指譯文通常按照譯入語的習慣用法。Laviosa（1998）將 Baker 的理論進一步驗證並發揚光大。Baker、Laviosa 以及其他學者採用兩個在文類分布比例接近但語言不同的語料庫，作為可比語料庫（comparable corpora）來研究譯文與非譯文兩者量化的差異，奠定了語料庫翻譯學的基礎。

一般而言，語料庫翻譯學的研究通常需要三種語料庫，分別是雙語平行語料庫、譯文語料庫以及非譯文語料庫，後兩者如果是根據同樣的原則按照主題或文類固定比例採樣，則可形成可比語料庫。我國由於與國際接軌，在語料庫的建設與應用研究起步相當早。1980 年代晚期，美國 IBM 與 AT&T 兩家公司的研究人員，開始利用雙語平行語料庫來發展不需要雙語辭典和規律的統計式機器翻譯系統。當時任教清華大學電機系的蘇克毅教授已經開始利用語料庫發展機器翻譯系統。1990 年初期，由中研院語言所黃居仁教授以及資訊所陳克健教授領導的團隊開始建構中文語料庫，至今中研院的漢語平衡語料庫已經有一千萬詞的規模，無論是對語言學的研究，或是中文譯文與一般非譯文的中文研究，漢語平衡語料庫都是不可或缺的資源。到了 1990 年代中期，包括臺灣大學陳信希教授和清華大學張俊盛教授，都有從事英漢平行語料庫的研究，他們研究的主題是如何利用演算法和機率，自動從英漢平行語料庫中取得句對應與詞對應的機率。這些議題雖然屬於計算語言學的範疇，但研究產生的句對應平行語料庫，無論對於翻譯教學及實務，或是研究譯者風格等議題，都是一項利器。

雖然我國在語料庫的建設與相關技術的研究早在 1990 年代初期即展開，但語料庫翻譯學引介進入國內卻是近十幾年的事。2000 年以後，臺灣及中國大陸才開始有較多學者注意到歐美語料庫翻譯學的進展。中國大陸翻譯語料庫的建設稍晚於臺灣，但在語料庫及人才培育雙管齊下之下，近年無論就翻譯語料庫的建設，或是相關著作都已經後來居上。執教於曲阜師範大學的夏雲教授便是大陸語料庫翻譯學の後起之秀，本

文要介紹的是夏雲教授根據她的博士論文改寫而成的英文專書，該書的中文題目是《翻譯學中的範化：20世紀英翻中小說歷時語料庫的研究》。

本書共有八章：第一章是概論；第二章是語料庫翻譯學研究的方法論；第三章是歷時語料庫的建立；第四章是詞彙的範化；第五章是搭配語的範化；第六章是語法的規範化；第七章是社會及文化環境與範式的關係；第八章是結論。以下摘錄各章節的重點。

第一章首先介紹翻譯學中跟範化相關的理論，以及為何要採取歷時語料庫的方法來進行翻譯學的研究。作者開宗明義介紹語料庫是什麼，以及語料庫翻譯學的發展沿革，並且指出語料庫翻譯學的哲學基礎源於西方哲學的理性主義對於自然規律的追求。從翻譯學的角度來看，語料庫翻譯學源於1990年代以來強調實證的描述學派，該學派的基本想法是，無論是一般的譯文或翻譯的過程，其本質上都存在某些趨勢。本身受過紮實語言學訓練的Baker在比較譯文與非譯文之後，發現了一些普遍性的特徵，Baker（1993）提出翻譯共性（translation universals）的理論。這些翻譯的共性是指與一般的非譯文相比，譯文在用詞和結構傾向較簡單（簡化），儘量語意清楚避免語意模糊（顯化），且在語言的用法會儘量符合一般譯入語的習慣用語（範化）。Baker（1996）對於範化的定義是「遵守譯入語的規律或習慣的一種趨勢」。本書作者提到，以跨越不同時代實證研究範化有幾個目的，這麼做可以研究不同時代背景、社會以及文化對翻譯活動的重要性。同時可以結合共時和歷時兩種觀點，以單語可比語料庫和平行語料庫來修正過去提出的研究方法。為了研究不同時代語言的習慣用法，作者建立同一本小說但不同時代的兩種譯本作為語料庫。

第二章介紹語料庫翻譯學研究的理論基礎，包括描述翻譯學、多元系統、語料庫語言學、語料庫翻譯學的研究方法及發展、譯文特徵、譯者風格、翻譯學中有關範化的實證研究等主題。作者引用Li與Zhang（2010）的研究，將語料庫翻譯學研究分成五個步驟，分別是（1）決定研究問題；（2）建立語料庫；（3）使用工具取得統計；（4）報告

統計的結果；（5）解釋統計結果所代表的意義。

第三章討論如何建立及利用歷時語料庫，首先提到編纂語料庫需要考量語料庫的類型、代表性、各種文類的平衡、語料庫的大小和取樣、可比語料庫等。接下來介紹歷時語料庫的設計、建立的目的地和類型、結構、文章的收集和取樣、語料的處理（包括中英文的句對應、語料的切分與詞性標注、分析工具）。作者總共建立了四個語料庫，分別是（1）30 到 40 年代的中文；（2）30 到 40 年代翻譯自英文小說的中文譯文；（3）1989 至 1993 年的中文；（4）1988 年至今翻譯自英文小說的中文譯文。其中（2）和（4）是相同的 14 本英文小說在不同時期的譯文。作者之所以選擇小說而沒有選擇其它文類，是在於早期的譯文以小說為主。這 14 本英文小說的譯者大多是單一譯者，有的則有兩位譯者。由於譯者沒有重複，兩個時期的譯者加起來多達 30 人，譯者多的好處是譯文可以反應出該時期的語言特色。為了方便研究不同時期的譯者如何翻譯原文，作者將（2）和（4）的譯文語料庫與原文的句子對齊，形成英漢平行語料庫。

第四章討論詞彙的範化，基本作法是比較同類型的譯文和非譯文的詞彙特徵。這些特徵包括各種詞性標記出現的次數，例如名詞、代名詞、高頻詞所佔的比例，詞與詞綴所組成的詞是否具有語意組合性等。作者的研究顯示，就詞性的分布而言，譯文和非譯文相當類似，最主要的差別在於中文譯文中普通名詞的比例顯著低於非譯文，而代名詞的比例則顯著較高。作者觀察中英對照的平行語料後發現，譯文中普通名詞的比例較低的原因，是由於英文原文有相當多的名物化名詞，而譯者在譯成中文時卻選擇動詞或形容詞。至於代名詞的比例，由於中文很少用非人稱的代名詞，因此代名詞在英文原文的比例遠高於中文。至於中文譯文的代名詞比例則顯著高於一般的中文，作者認為有兩個可能的原因：（1）受到英文原文的影響；（2）翻譯過程中顯化。根據 Lavisosa（1998）的研究，高頻詞在譯文中的比例較非譯文高。作者假設譯文與非譯文高頻詞重複率越高，越能顯示非譯文的代表性以及譯文的範化。作者比較

30 到 40 年代的譯文與非譯文語料，以及 80 年代晚期的譯文與非譯文之後，發現高頻詞重複率在前者較高。這顯示早期譯文的範化比晚期的譯文更明顯。語言學家早已發現古時中文的詞長幾乎都是單音節詞或是雙音節詞，但受到西方語言的影響，多音節詞的比率逐漸提高。作者發現 30 到 40 年代的中文非譯文語料庫中的單音節詞比率，在四個語料庫中最高。同時兩個時代譯文語料庫中，多音節詞的比率也都比非譯文語料庫高。而比較兩個不同時代譯文語料庫則會發現，早期的譯文語料庫單音節詞的比率較高，換言之，早期的譯文比較接近同時期的非譯文。最後作者比較「化」、「性」、「度」、「家」、「者」這五個詞綴在四個語料庫出現的比率，發現這五個詞綴在早期的譯文語料庫出現比率較高，明顯受到英文原文的影響。因為中文歐化已經根深蒂固，整體而言 80 年代晚期的譯文與非譯文中，這五個詞綴的比例並沒有比較高。

第五章探討搭配語的範化，首先作者介紹搭配語的定義，是指在三到五個詞範圍內經常一起出現的詞彙組合，並根據文獻歸納出語料庫翻譯學有關於搭配語的研究有兩種。第一種是利用平行語料中的原文和譯文，來研究兩個語言間的轉換 (shift) 以及策略；另一種則是利用相同的語言中譯文和非譯文的可比語料，來分析譯文在使用搭配語時，是否有某些普遍性的特徵。作者找出 12 個關鍵詞 (其中一半是動詞，一半是名詞)，這些詞符合下列四個條件 (1) 在四個語料庫中的次數都大於 100 次；(2) 在四個語料庫中出現的次數儘可能接近；(3) 以互見訊息 (mutual information) 大於 4 作為搭配語認定的標準；(4) 去掉次數小於三次或僅出現在三篇文章以內的搭配語，以減少作者、譯者或是主題的影響。本書作者以搭配語的數量與關鍵詞比率作為比較的標準，並且發現在 12 個關鍵詞中，其中有 9 個詞在早期譯文語料庫中有較少的搭配語。為了計算搭配語的多樣性，作者再比較四個語料庫中「的」或「底」後面接的名詞各有多少種。「的」或「底」是相當高頻的功能詞，有越多種的名詞搭配語，表示這個語料庫搭配語的多樣性越高。多樣性計算的方式是搭配語種類的數量除以總共出現幾次。重複率則與

多樣性呈反比。在四個語料庫中，早期譯文的重複率比同時期的非譯文低，同時標準差也比較低，這表示搭配語的多樣性較高，也可視為範化的反證。最後作者從中英平行關鍵詞檢索程式，來檢視所有「的」或「底」後面接的名詞的例子。作者發現雖然有不少翻譯符合中文的習慣，但也有一些翻譯直接根據英文原文字面上的意義翻譯，不符合中文的習慣。這些例子包括“humble pie”「卑微的餅」、「a careful outline」「細心的輪廓」、「a faint and bare existence」「微末底生存」、「a bright fire」「光輝底爐火」。作者認為這些都屬於去除範化（denormalization）的例子，也就是使用新的或是特殊的搭配語。

第六章討論語法的範化，作者首先引用文獻，說明中文與英文在語法上的不同之處，在於英文是形合的語言，句子內部的邏輯和語義關係依靠詞性和結構。而中文是義合的語言，句子的內部邏輯和語義關係並不是依靠詞性，而是靠語境來推論。然而由於與西方的接觸，語言歐化的影響造成翻譯作品中連接詞的比例逐漸增多，這一點類似形合的語言。本章利用統計的方法探討連接詞、介詞、被動結構以及句長在不同時期的範化，作者以 log likelihood test 測試發現，在兩個不同的時期，譯文中無論就連接詞的種類或連接詞的比例，都顯著高於同一時期的非譯文，這顯示早期的譯文因為受到英文原文的影響，比較接近形合語言，且早期的譯文中連接詞的比例也顯著高於較晚期的譯文。另外早期的譯文與非譯文無論就連接詞的種類或頻率，都高於較晚期的譯文與非譯文。從譯文與非譯文兩者共有的詞數與比率來看，早期的譯文與非譯文無論就共有的詞數與比率，都低於較晚期的譯文與非譯文，這顯示早期的譯文有較多的連接詞無法在同時期的非譯文找到，其變異性較大。而較晚期譯文的連接詞與同時期的非譯文較接近，範化較明顯。介詞的比例也是譯文高於同時期的非譯文。中文的被動句型有「被」、「為……所」、「給」、「叫」、「讓」等幾種。以比例來看，早期的譯文、較晚期的譯文與較晚期的非譯文三者差異不太，但以早期的譯文出現最多，而同時期的非譯文最少，只有略超過一半。由此可以看出早期譯文

中的被動句受到英文原文的影響，而與同時期的非譯文差異最大，範化的影響最低。就平均句長而言，早期的譯文也是四個語料庫中最長，這也是歐化影響的另一個例證。

第七章討論社會及文化環境與範式的關係，作者提出範化與去除範化兩者是共存的，特殊的社會及文化環境決定了翻譯時要導入多少的異質元素以及文化的接受程度，20世紀早期的譯者之所以引入大量的歐化語言，主要是因為他們將這些西方語言視為具有革命性，較精確且有邏輯，一如西方的思想。這種過度偏向西方語言且生硬的翻譯方式，已隨著時間不斷自我調整後逐漸消失。

第八章是結論，作者總結出研究的發現以及對理論隱含的意義。從本研究可以得知，先前對於範化的假設有必要進行擴充和修正。本研究支持 Kenny (1998) 提出的看法，認為非典型或非範化的特徵與範化一起出現，這是由於譯者在翻譯的過程當中面臨原文和譯文兩種文化的拉鋸，以及「信」與「達」之間的衝突，造成範化及去除範化兩種同時共存的競爭。本書也指出某些社會文化因素，例如文化強弱關係、翻譯理論的發展、語言的態度等，都有可能對翻譯行為的規範產生影響。作者認為就方法而論，本書結合了兩個不同時代的譯文語料庫及相同時代的非譯文語料庫，也採用了平行語料庫雙語的例句，並從譯入語以及歷時的角度切入，為範化提供了一個較為全面的面貌。至於研究的限制，作者認為主要是語料庫仍然不夠大，無法顧及作家的代表性與平衡性。一個更大型的語料庫可以包含更多不同文類的文章，以及更多歷時的語料庫，如果有更大型的平行語料庫及可比語料庫，則更能歸納及推廣結論。另外作者也承認，雖然分詞以及詞性標記都經過人工檢查，但是中間仍然可能存在錯誤。

本書文字流暢，文獻回顧相當詳細，研究方法嚴謹，論述清楚有力，無論就議題、內容、所使用的語料庫資源與技術，都有許多地方值得借鏡。由於本書內容包含中文自動分詞、中文自動詞性標注、中文自動分句、半自動中英雙語句對應、LLR 統計顯著性檢定、以及相關語言學

和翻譯學理論的術語和概念，沒有語料庫翻譯學背景的人閱讀此書，恐怕會感到吃力。本書沒有指出的一個問題是中文句子如何辨識，由於句號、問號、驚嘆號等標點符號不能完全表示中文句子的界限，書中有關平均句長的計算不無疑問。此外，以互見訊息（mutual information）大於 4 作為搭配語認定的標準，也有一些爭議，原因是互見訊息大於 4 的計算方式是根據作者的經驗得到，並非理論的值，且只使用互見訊息會使得高頻詞的搭配語偏低，一般會用搭配 t-score 一起考慮。

筆者想補充的是，語料庫翻譯學的哲學基礎除了理性主義也包含實證主義。Baker 的理論或多或少受到現代語言學大師杭士基（Chomsky）的語言共性理論所影響。杭士基認為，世界上所有的語言儘管表面有許多的差異，但在抽象的層次具有共同的特徵。Baker 翻譯共性理論明顯受到杭士基語言共性理論的啟發。但不同於杭士基透過衍生學派（generative grammar）各種極為抽象的理論來證明語言共性，Baker 依據語料庫的詞頻、搭配語和句長等統計來證明翻譯共性。換言之，Baker 的方法論是基於實證的語料。本書作者的研究方法和理論假設基本上與 Baker 相同，然而必須指出的是，語料庫翻譯學的發展至今也不過 20 多年的歷史，翻譯共性屬於 Baker 早期至中期的研究議題，2000 年以後，Baker 的研究興趣逐漸轉移到別的議題。雖然還是有不少學者致力於有關翻譯共性的研究，但許多學者已經聚焦在如何利用語料庫研究譯者的風格等其它議題。無論是翻譯共性，或是譯者風格等議題，都需要有語料庫及相關的技術來支撐。有志於語料庫翻譯學研究的學者，如果沒有現成的語料庫可以用，就必須瞭解如何利用各類的工具自行建構語料庫，包括中文自動分詞、中文自動詞性標注、中文自動分句、半自動中英雙語句對應、LLR 統計顯著性檢定，並且熟悉相關語言學和翻譯學理論的術語和概念，才能從事相關的研究。我們相信語料庫計算語言學的理论以及所開發的各式工具，將影響語料庫翻譯學未來的走向，特別是關於譯者風格等議題，在新的語料庫工具與方法輔助下，有可能使得語料庫翻譯學更趨向科技整合的人文計算（humanities computing）



學科，而非傳統的翻譯學。如同本書的作者使用許多語料庫技術以及統計工具作為研究的工具，可以預期未來將有學者使用資料探勘（data mining）和機器學習（machine learning）的理論及工具程式，來處理語料庫翻譯學的研究議題。

## 參考文獻

### 中文文獻

王克非、黃立波（2008）。語料庫翻譯學十五年。中國外語，5（6），9-15。

【Wang, K. F., & Huang, L. B. (2008). Corpus-based translation studies: Progress in recent 15 years. *Foreign Languages in China*, 5(6), 9-15.】

### 英文文獻

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233-250). Amsterdam, Netherland: John Benjamin.

Baker, M. (1996). Corpus-based translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-243.

Kenny, D. (1998). Creatures of habit: What translators usually do with words. *Meta*, 3(4), 515-523.

Laviosa, S. (1998). The corpus-based approach: A new paradigm in translation studies. *Meta*, 43(4), 474-479.

Li, D., & Zhang, C. (2010). Sense-making in corpus-assisted translation research: A review of corpus-assisted translation research in China. In R. Xiao (Ed.), *Using corpora in contrastive and translation studies* (pp. 235-255). Newcastle, England: Cambridge Scholars.