

華語學習者句式使用情況分析*

張莉萍

國立台灣師範大學

摘要

本研究藉由以任務為導向的電腦寫作考試語料來分析量化不同程度華語學習者的句式使用情況，目的是為了找出不同能力學習者的關鍵語言特徵，以作為評量時的參照指標。研究方法包括母語與二語語料庫之間的對比，以及偏誤分析（正誤用例比例）的方法，以找出學習者過度使用或少用哪些語言特徵、不同能力學習者的習得情況。對於是否為關鍵性特徵的判定則以 Hawkins and Buttery (2010) 對關鍵特徵的判定標準為主，以統計上的差異是否具有顯著性為依據。本文以「把字句、被字句、連字句」為例，觀察分析這些句式的使用頻率、偏誤率、次結構使用情況。這些資訊將有助於華語教學大綱的制定和教學材料的編寫，為學習、教學與評量設計提供科學的根據。

關鍵詞：華語、句式、關鍵特徵、學習者語料庫、偏誤分析

1. 研究背景與動機

近年來，台灣華語教學或測驗單位以歐洲語言共同參考架構（Common European Framework of Reference for Languages: Learning, Teaching, Assessment）（以下簡稱 CEFR）（Council of Europe 2001）為教學、教材或評量分級架構者已有增多趨勢（藍珮君 2007; 謝妙玲、林雅惠 2008; 蔡雅薰 2009），不過，由於這個架構不是針對特定語言所建立，因此沒有為特定語言建立充分而詳細的語法、詞彙等語言知識內容，但對於教學或評量而言，除了抽象能力指標的描述外，要能達到這些能力指標所需具備的語言知識是不可或缺的學習或評量元素。因此擬定針對各等級語言能力的語言知識內容，以有效做為教學與評量

* 本研究感謝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」資助，並感謝兩名匿名審查人寶貴意見。

的標的，有其必要性。否則即使教學系統或評量工具宣稱對應於 CEFR，可能在語言內容或知識上的設定仍存在不小落差（張莉萍 2012）。

歐洲理事會（Council of Europe 2005）爲了將這些抽象的能力指標落實在特定語言的教學與測試中，已經進一步擬出等級描述參考（Reference Level Descriptions for National and Regional Languages (RLD)）的說明報告，在這份報告中指出等級描述的目的，是爲了清楚地把共同架構中的指標落實在特定語言不同程度的使用者身上，藉著該語言的內涵和應該達到的語言表現來描述或改述這些指標。英國的劍橋大學與其他合作機構所進行的英語概貌計畫（English Profile Program），就是爲了產出一套英語等級描述參考而進行的跨領域共同合作的長期研究計畫，以能更精確地在教學或評量上使用 CEFR 的能做描述（can-do statements），現階段已經產出不少成果，包括對應於 A1-B2 等級的詞彙內容（Capel 2010），以及各等級的關鍵語言特徵（criterial features）（Hawkins and Buttery 2010）。

張莉萍（2012）的研究已經爲對應 CEFR 能力等級提出了詞彙量的建議，本研究則是擴展到詞彙之外，觀察分析學習者在句式使用上的表現，期望透過量化與質化的分析，嘗試建立有區辨度的語法、句式等語言知識內容。

舉例來說，American Council on the Teaching of Foreign Languages（ACTFL）所制訂的中文能力綱要（Chinese Proficiency Guidelines）（ACTFL 1987）中針對每一個等級所舉的代表性例子，可以視爲關鍵特徵，撰寫者以列舉語言特徵的方式來說明不同能力學習者達到的語言能力內涵，這些語言特徵是針對某一等級特別突出的特徵，以下是書寫能力中上程度（intermediate - high level）的中文綱要內容¹。

……能稍微準確地透過時間詞的使用，而不用語助詞來表達時間的框架，假如過去貌標誌「了」的掌握，能對應於寫作者母語的過去式，他便較能掌握。基本上寫作者能掌握語法句型及全數的言談結構，除了一些較特殊的 S-V-O 結構，如主題顯著的賓語提前（工作，他還沒做完。）、結果/方向結構（搬進來）、及關係子句修飾語（我們去年買的+名詞），此時仍會出現常有的錯誤。「把」字結構、「是……的」句型仍會使用錯誤。已能使用口語為基礎的句性副詞和連接詞，如「要不然、雖然……可是」。已出現連續的言談結構。（翻譯自 ACTFL 1987:485）

¹ 翻譯引自 2001 年鄧守信教授指導臺灣師範大學華語文教學研究所楊忠諺等研究生翻譯〈美國外語協會中文程度綱要〉文件。該文件未出版。

此段能力敘述中，撰寫者藉著學習者對「了、時間詞、把字句、是...的、動補結構、賓語前提、關係子句修飾語、要不然、雖然、可是」等等的使用來說明語言使用者在這個階段的語言表現。

我們嘗試要做的，即是找出類似中文能力綱要描述中的這些特徵，描述出對應於 CEFR 能力指標的中文特徵。根據 CEFR 指標研究的創立學者之一 Brian North 的建議²，CEFR 和 ACTFL 的等級對應合理的假設是：C2 是 Distinguished，C1 是 Superior，B2 是 Advanced-mid，B1 是 Intermediate-high。ACTFL 為外語所制訂的能力綱要，最早源自美國外事學院（Foreign Service Institute, FSI）所製作的口語能力 5 級量表（0-5），後又稱 ILR（Interagency Language Roundtable）量表。我們將 ACTFL、FSI/ILR、CEFR 等級關係簡示如表 1³，以利讀者參照。

表 1 CEFR, FSI/ILR, ACTFL 等級對照

CEFR	FSI/ILR	ACTFL
A1	0/0+	Novice-low, Novice-mid, Novice-high
A2	1	Intermediate-low, intermediate-mid
B1	1+	Intermediate-high
B2	2/2+	Advanced-mid
C1	3/3+	Superior
C2	4/4+	Distinguished

在這個報告中，我們主要分析把字句、被字句和連字句，這三個句式在鄧守信（2009）《對外漢語教學語法》中都是屬於結構複雜度高的句式。

2. 研究方法

對於不同能力學習者的語言表現，我們希望能以實證方法來分析，儘量避免僅止於專家經驗式或直覺式的判斷。因此本文主要是基於語料庫為本的研究。

2.1 對比中介語分析方法

此研究主要採用對比中介語分析方法（Granger 1998:12），透過本國人語料庫（L1）和學習者語料庫（L2）的比較，找出學習者少用或過度使用的語言特徵。L1 語料庫主要採用中央研究院平衡語料庫（中央研究院資訊科學研究所等），約五百萬詞；L2 語料庫則是 TOCFL 學習者語料庫（張莉萍 2013）。TOCFL（華語文能力測驗）是電腦考試形式，考生直接於線上輸入文字。該語

² 根據 2006 年 3 月 Brian North 參加日本外國語大學國際研討會，於現場口頭回答提問時所提出的建議。

³ A1 與 A2 的對照則是參考 American University Center of Provence 所制訂的對應表。來源：http://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages。

料庫蒐集 2006 年至今參加 TOCFL 考生的作文，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，關於不同能力學習者的語料量如表 2 所示。

表 2 TOCFL 學習者語料庫語料分佈情況

能力等級 (CEFR)	主題數 (Topic)	詞數 (Token)
A2	23	120,754
B1	26	445,997
B2	17	306,303
C1/C2	64	263,432
小計	130	1,136,486

2.2 偏誤的認定與計算

首先在量化時，會遇到偏誤認定與計算的問題。例如，拿「見面、進步、聊天、說明、保證」等詞語來說明，這類動詞的賓語（對象）不能直接出現在動詞後面，但 B1 學習者會寫出「我見面一個美女」、「以後來報告我一下」這樣的句子，如果以一般偏誤率的的算法（以錯誤的筆數除以詞語出現的筆數），平均偏誤率僅 3.6% 左右，這樣看來，似乎顯示學習者已經掌握了這些詞語。其實這種算法的偏誤率並不能客觀地呈現學習者對這些詞語實際掌握的能力，因為學習者使用這些詞語，很多時候並不牽涉及物性的表達，例如，「聽了她的說明以後，…」、「我們終於見面了」，分析者無法從這些用例得知學習者對及物性用法的掌握度。因此我們的作法如下，以「見面」為例，實際統計每個等級對及物性用法的正誤比例。

表 3 「見面」及物性使用情況

	出現筆數	使用「跟 X」於前	使用對象於「見面」後	偏誤率%
A2	118	29	9	23.68
B1	653	177	35	16.51
B2	76	22	3	12.00
C	19	5	0	0

從表 3 得知 B2 語料中出現「見面」的筆數是 76 筆，其中正確使用「跟 X」於「見面」前的位置有 22 筆，另有三筆錯誤用法是直接將見面的對象放在「見面」這個動詞後，顯示即使 B2 學習者仍有 12%（3/25）的偏誤率，表示仍未習得。「見面」這個例子相當具代表性，因為四個等級中都有用例，而且也看得出來偏誤率隨著學習者能力提升而下降。但是並非每個詞語或語法點都能找到完美的分佈，例如，「結婚」這個詞語在四個等級中出現了 360 筆，其中

26 筆牽涉到及物用法的，只有一筆使用錯誤（在 A2 等級中，將結婚對象直接放在「結婚」後），僅能說 A2 學習者使用了 3 筆中，有一筆錯誤，偏誤率是 33.3%。B1 以上學習者在使用這個詞語時，沒有問題。那麼究竟及物性是不是普遍的區辨性特徵，還是個別詞語的特徵？這個特徵似乎需要更多統計數據才能驗證。如果從這兩個詞的使用情況來看，僅能說明「非常規賓語位置」這個特徵，A2 學習者還無法掌握。

除了上述偏誤計算方式的問題，第二個量化的問題是，在觀察學習者句式表現時，我們發現前人對偏誤判定的相關研究，除了「當不當用」這個句式的判定外，還針對該結構的內部成分，採用遺漏、誤加、誤代、錯序來統計分析偏誤（肖奚強 2009），但這樣所呈現的正誤比例不見得是這個句式的習得情況。因為如果連句式的內部構成成分也都要學習者完全無誤的表達，這個要求是過高的。判定句式偏誤用例的原則應該不同於判定詞語偏誤。詞語偏誤的確可以大致清楚地以上述四個偏誤類型來判定；然而判定句式的偏誤原則則複雜得多，不同階段學習者能搭配句式使用的詞語或與該句式相關的其他語法掌握度未必同步，可能一個句子出現多處偏誤，也可能僅是周邊一個微不足道的偏誤。例如，肖奚強（2009:31）將「…把我送到去醫院」計入「把字句」誤加類的偏誤用例中，認為「去」為誤加或「去」應該移至句末，雖然精確的表達是如此，然而就學習者是否會自然地產出把字句、是否能掌握基本句式，而且在正確時機點使用而言，顯然學習者是做到了。因此我們並不會將類似情況判定為偏誤用例。由於每個句式偏誤判斷的實務過程太瑣細，本文略過這個部分。

2.3 句式習得的判定標準

對於學習者習得與否的判定，本研究採用傳統習得理論中的習得標準（Brown 1973），即是正確率達 90% 以上可以算是習得。依據這個標準，表 3 的數據可以顯示 B2 學習者仍然沒有習得「見面」的用法。然而，對於學習者是否掌握了句式的次結構，則較難採取這個方法。例如，可以從語料中看到 A2（初級）學習者使用了近十次「把中文學好」這樣的句式，但學習者不可能在 A2 這個階段掌握了把字句。

因此除了使用正確率的標準外，還有一個判斷標準，即 Pienemann（1998）的 *emergence criterion*，張燕吟（2003）將 *emergence criterion* 翻譯為「初現率標準」，本文採曹逢甫教授建議改為「初現標準」。「初現標準」是從語言處理的角度來說，「初現」代表在某一時間點，某種技能已經在原則上得到掌握。而這個概念與我們希望尋找出在不同能力等級學習者的關鍵特徵很接近。至於「初現」結構的定義則是「第一次有系統、能產性地使用」，因此我們判定「初現」的原則主要有兩個，一是有系統性的，一是非公式化的。像

A2 學習者說出「把中文學好」這個語言形式，不代表他們真能使用帶有補語的把字句結構。

2.4 關鍵特徵的標準

關鍵特徵之所以關鍵，是指這項特徵會隨著使用者不同的能力而有不同表現，因此這樣的特徵可以用來當作測量不同能力等級的學習者的依據（Salamoura and Saville 2009）。

至於各等級間表現的差異是不是具有顯著性，我們採用 Hawkins and Buttery（2010）關鍵特徵的作法，兩級間的差異至少要達到 29%，才是顯著的。以表 3 為例，可以知道偏誤率的顯著差異呈現在 A2、B1 之間，B2 和 C 之間。B1 和 B2 之間雖然有差異，不過未達顯著標準。也就是說，「見面」可以用來區辨三個群體——[A2]、[B1、B2]、[C]。

3. 華語常用句式分析

在這一節，我們的重點放在把字句、被字句、連字句是否為關鍵特徵，也就是說，在不同能力學習者間的使用率或偏誤率是否有顯著差異，進而描述學習者語料庫中相關句式次類結構的使用情形。

關於這三個句式的相關語法、語義或語用分析不少，不過在下面幾個小節，我們僅針對同樣是採用語料庫方法所做的研究提出討論。為了客觀呈現學習者的習得情況，研究將呈現不同語料庫所做出來的結果，除了台灣的 TOCFL 學習者語料庫外，還有大陸的 HSK 動態作文語料庫（張寶林、崔希亮、任傑 2004）、南京師範大學外國學生漢語中介語偏誤信息語料庫（肖奚強 2009）。

3.1 把字句

3.1.1 使用率

關於學習者使用把字句的部分，我們拿都採用語料庫方法所做的研究——肖奚強（2009）、張寶林（2010）和張莉萍（2013）數據來看（見表 4），肖奚強（2009）的研究方法是，統計了中介語語料（L2）和本國人語料（L1）各 90 萬字的用例，結果是 L1 使用了 1,285 次，L2 使用了 728 次，從而得出 L1 的使用頻率是 L2 的 1.77 倍（肖奚強 2009:27），他的結論是「外國學生確實存在把字句使用的迴避現象」。然而張寶林與張莉萍的研究結果並沒有顯示學習者有這個傾向。

表 4 把字句的使用率⁴

把字句使用率	張莉萍 (2013)	張寶林 (2010)	肖奚強 (2009)
L1 語料庫	0.144%	0.076%	1,285 次
L2 語料庫	0.138%	0.092%	728 次

在這三個分析中，所採用的語料庫大小，不論 L1 或 L2 語料庫，張寶林 (2010) 所採用的語料量都遠遠超過其他兩者；肖奚強 (2009) 則是這三個中，數量最小的。就語料的平衡性而言，張寶林 (2010) L1 採人民日報都是新聞語體的語料，L2 是高級程度語料 (HSK 寫作考試語料)；肖書 L1 語料都是文學性文章，L2 則是初、中、高程度各 30 萬字；張莉萍 (2013) 所採用的 L1 語料是中研院平衡語料庫，L2 則是各種程度都有 (TOCFL 寫作考試語料)。就整體而言，張寶林 (2010) 與 TOCFL 語料庫的代表性應該是較佳的，雖然還不是最理想的。綜合兩者 L2 與 L1 語料庫資料顯示，L2 使用把字句的頻率與 L1 並沒有顯著差異。

雖然學習者整體使用把字句的頻率與母語者沒有顯著差異，然而我們觀察不同能力學習者之間的使用情況，發現是有顯著差異的。如表 5 所示，使用率基本上隨著語言能力提高而增加，C 級使用率雖然沒有上升，但與 L1 的 0.144% 也相差無幾。如果從 A2、B1 學習者使用率低於平均來看，一般教師或專家學者認為學習者迴避使用把字句的「感覺」也就沒錯，而前述 HSK 語料庫所收集的是高程度考生作文語料，L2 的使用率 (0.092%) 還稍高於 L1 (約 0.076%)。

表 5 把字句在 TOCFL 學習者語料庫中分佈情況

級別	使用次數	使用率 (%)	偏誤率%
A2	106	0.0878	43.40
B1	544	0.1220	18.08
B2	643	0.2099	11.20
C	353	0.1340	7.08

可見談使用率，想要探討學習者是否多用或少用某結構或語法點，精確的作法必須分階段來觀察分析，因為學習者的使用情況會因為能力不同而不同。就把字句的使用情況，我們大致可以看出學習者在從 B1 過渡到 B2 這個階段，使用頻率已經趨向目標語人士。

⁴ 我們發現大陸學者使用頻率的算法，多是以出現次數除以總字數來計算，台灣學者分母多以總詞數來計算。不過，並不影響我們的觀察結果，因為我們不做兩岸頻率的對比，而是純粹看該地目標語人士與學習者的使用情況。如果依據 Chen et al. (1993:83) 指出中文平均詞長是 1.494 來換算 (也就是以總詞數而非字數做分母) 張寶林文章中學習者的使用率約為 0.138%，與 TOCFL 學習者的頻率非常接近。

3.1.2 偏誤率

肖奚強（2009）統計偏誤的方式包括對把字內部結構的成分進行遺漏、誤加、誤代等統計，平均偏誤率為 25.94%⁵；張寶林（2010）則是考察了 HSK 語料庫中 3,682 個把字句中，得到偏誤句 461 句，佔把字句總數的 12.52%，認為大多數的學習者是可以正確使用把字句的，以往大家誇大了把字句的難度。我們則是觀察分析了 TOCFL 語料庫中 1,646 個把字句，表 5 已經顯示了每個等級的偏誤率，平均偏誤率是 19.94%，B2 階段的偏誤率接近 HSK 高等寫作考試考生的數據，同時也可以看出這個偏高的偏誤率是因為 A2 階段的極高偏誤率所致。從這三個研究的數據中，張寶林的數據僅能說明高程度學習者習得把字句的情況，肖奚強（2009）的數據則因為偏誤的認定不單純，無法呈現把字句真實習得的面貌，TOCFL 的數據呈現較全面性，每個等級偏誤率的差異也都達到了顯著標準。從這個觀點看，把字句毫無疑問可以做為區辨不同等級學習者的關鍵特徵。

3.1.3 次結構使用情形

我們觀察了 A2 語料庫中 106 個把字用例發現，偏誤率達 43.4%，也就是正確率連百分之六十都不到。學習者使用的次結構是結果補語和目標結構⁶，但因為偏誤率高，表示學習者無法掌握；使用「光桿動詞+了」結構的只有一筆「把…忘了」用例，可能與學習者習得的動詞有限有關；使用「V — V」也是有一筆「把電視看一看」，是錯誤用例。依據系統性和能產性這兩個原則，A2 學習者不具有使用把字句的能力，雖然可以看到學習者使用結果補語和目標結構，但都是較公式化的用法，例如，「把中文學好」這樣的用例出現了近 10 次。

限於篇幅，在此就不詳述每一等級觀察到的現象，而嘗試以下面的方式來描述不同能力者把字句的使用情形，描寫的原則以學習者能做到的為主，不描述無法觀察到或學習者做不到的部分，在前一級已經出現的語法結構不再重複描述。

A2：利用有限的詞彙使用固定的把字句式，典型的例子：把中文學好、把…拿出來、把…送到機場/放在桌上。

B1：能使用基本把字句型，包括「把」加結果補語、目標結構（在/到/給）、得字補語、「V 作/做/成」、「光桿動詞+了」、「V — V」、「V+動

⁵ 這個偏誤率是我們就書中第 25 頁表 1.4 中的數據，自行計算得出 193（誤例）/744（誤例+正例）= 0.2594。

⁶ 為了簡化表達，我們所指的結果補語包括了一般學者所指的趨向補語，目標結構則指「把…V 到/在/給」。肖奚強（2009）所指的情態補語，我們則以「得字補語」表示。

量結構」、「V+間接賓語」。也能將把字句和兼語句一起使用。

B2：出現「V 爲」結構，能擴展基本把字句式而自如產出非基式的句子，並能結合其他句式一起使用，例如，連動句「把寵物租回家玩幾天」、「把它送到二樓給你吃」，被字句「他們想盡辦法把我害得差點兒被學校開除」。

C：出現「V 向、V 於、V 化」結構，如「把…公開化」。能更穩定地使用各類把字句（準確率在 90% 以上）；能與其他句型共同使用，例如，在連動句中使用，如「他們負責把東西從車上搬下來」；能靈活運用把字句，例如，在動詞前加各式狀語，如「導演的義務就是把故事在三個鐘頭以內敘述完」、「把還活的蝾螺直接放在鐵板上」。

在描寫學習者能力時，如何判定這是一個次結構，是否要予以描述？我們還沒有很好的標準。例如，「V 爲」應該和「V 作/做/成」是同類結構，只因爲動詞補語的形式不同而分別描繪，似乎有一致性的問題。不過，考量到其實把字句次結構中，大部分都是屬於補語範疇，一般教學者或研究者在談論把字結構時，也往往把結果補語或趨向補語區分開來談，或區分名量補語和數量補語；現階段我們還是以學習者的實際產出做事實描述，否則又回到以往經驗式描述的方法。

3.2 被字句

3.2.1 使用率

肖奚強（2009）基本上將被字句依照施事者（agent）是否出現以及動詞後面接的成分⁷，分爲六類：

I	N1 + 被 + N2 + V	他被媽媽打了
II	N1 + 被 + N2 + V + N3	他被當地群眾稱爲農民詩人
III	N1 + 被 + N2 + V + C(C 表補語)	她不希望被這男人搞得暈頭轉向
IV	N1 + 被 + V	這個學校爲什麼不被承認
V	N1 + 被 + V + N2	有的詩歌已經被譜了曲
VI	N1 + 被 + V + C	什麼藝術品都被搬運一空

統計出的 L1 和 L2 在各自 90 萬字的語料庫中使用次數分別是 1,208 和 261 次，得出學習者被字句的使用量嚴重不足的結論。「被字句」在 TOCFL 語料庫與平衡語料庫的使用頻率則分別是 0.096% 與 0.148%⁸。顯現雖然學習者有不

⁷ 肖書以賓語來稱呼「被」之後緊接的名詞組，爲避免混淆，我們以施事者或 N2 稱之。

⁸ 我們用 1,000 萬詞平衡語料庫的版本得出被字句的使用率爲 0.1427%，這個數據與 500 萬詞平衡語料庫的版本幾乎相同。

足使用的情況，但沒有肖奚強（2009）所得出的差距五倍之多，而是大概 1.5 倍左右。下表是不同等級使用率，可以看出 B2 階段的使用率(0.160%)即和 L1(0.148%)不相上下，這個發展與把字句相似。再次顯示，對比學習者與 L1 的使用率差異，不同能力學習者的表現會不同，分析時應該有針對性。

表 6 被字句在 TOCFL 學習者語料庫中的分佈

級別	使用次數	使用率 (%)	偏誤率%
A2	20	0.0166	45.00
B1	188	0.0422	15.96
B2	491	0.1603	7.74
C	435	0.1651	5.55

至於被字後面接施事者（N2）的情況是否如肖奚強（2009:41）調查結果所言，學習者傾向帶 N2 情況較多，與 L1 相反。我們統計 TOCFL 語料庫 1,134 筆被字句用例中，得知「被」後面緊接動詞的有 568 筆，顯示佔了總數的一半左右，另外，李婕、陳晨（2012）根據 HSK 語料庫所統計分析的結果，顯示緊接動詞的用例佔了總數的 64.72%，可以看出，並沒有如肖奚強所言，學習者使用被字句時傾向於帶 N2。由於 TOCFL 與 HSK 語料量都較肖書用例為多，結果應該較可信，本研究顯示學習者在被字句的使用上並沒有帶施事者的傾向。

爲了驗證能力越高學習者使用的傾向也越接近 L1，我們分別統計了各級學習者後接動詞的比例，結果顯示，A2, B1, B2, C 四個級別的百分比分別爲 25%、38.30%、50.71%、55.63%，可以清楚地看出學習者的使用情形隨著能力的提升而趨近目標語人士，也就是初級學習者一開始傾向帶施事者，到了 B2 階段，已經沒有這個傾向了。

3.2.2 偏誤率

如表 6 所示，可以清楚看出隨著能力提升，偏誤減少的趨勢，除了 B2 到 C 之間沒有顯著差異，其他各鄰近等級的差異是顯著的。依據 Hawkins and Buttery（2010）的標準，被字句可以區辨[A2]、[B1]、[B2、C]這三群學習者。從表 6 也可以看出，B2 階段學習者對被字句的使用已經趨向穩定，偏誤率在 10%以內，似乎對這個階段學習者而言，「被字句」掌握得比「把字句」好。而四個級數總平均偏誤率是 18.56%，大部分是受 A2 高偏誤率的影響。肖書的總平均偏誤率爲 18.77%（p.40，偏誤用例 49 例）；另根據 HSK 語料庫統計出的偏誤率都在 11-12%左右（桂孟秋 2011；李婕、陳晨 2012）。

3.2.3 次結構使用情形

我們從觀察學習者語料中，可以知道大部分研究者所提到的被字句型在 B1

階段都已經出現，TOCFL 語料庫還有肖奚強（2009）所未提及的「被…給/所 V…」句式。B2 以上學習者的語言表現與 B1 最大的不同，應該是在他們可以自如運用被字結構出現在句中的各個位置，而非單純地做爲主要謂語，以及因爲能力的提升，而能使用多元的動詞與「被」搭配。不過，我們也發現高能力者仍存在不少不該用被字句而用的情況，而這些情況通常是因爲詞語本身有「遭受義」，學習者可能是因爲語義關係而自動使用被字句，例如，「*希望沒有人再被（食物）中毒」、「*社會經驗少的人容易被歹徒上當」。這個情況應該可以避免，如果學習者意識到「中毒、上當」這些動詞屬於不及物動詞，不能使用在被字句中，就不至於有這類偏誤。只是現階段華語教材並沒有關於動詞及物性的標記，及物性標記在教學上實在有其必要性。以下是針對 A2-C 級學習者在被字句上的能力描述。

A2：能使用典型的、較公式化的「被字句」，例如，「被…罵了（一頓）」、「被…偷（走）了」。

B1：能使用基本的被字句型，動詞後面加補語的形式漸漸豐富，包括結果補語、趨向補語、目標結構補語。會使用「被…給/所 V…」句式。基本上，還是應用在簡單句中，帶施事者的頻率較不帶施事者的高；少數已能使用在連動句中，如「他怕被騙」。

B2：能掌握「N1+被+N2+V（爲/成/作）+N3」結構；能使用動詞後面得字補語結構。能自由運用被字結構在句中不同的位置，如主語位置「被人誤會的感覺總是不好受的」、賓語位置「一杯裝著被污染的水」。能和其他句式一起使用。能更穩定地使用被字結構（準確率在 90% 以上）。

C：出現「V 化」形式，例如，「被…半殖民地化、被邊緣化、被…同化」；與「被」搭配使用的動詞更多元。不出現 N2 的頻率較出現來得高。能穩定地使用被字句。

從上述描述大致可以看出，我們認知的「被字句」次結構標準同「把字句」類似。區分動詞後面的補語結構類型（結果、趨向、得字句等等）、必須出現的間接賓語句型、V+爲/成/作、V+化；不同於把字句的，則是被字句有文言殘留的「被…給/所 V…」結構以及「被」後面的名詞可以省略的結構。

3.3 連字句

3.3.1 使用率

這裡的連字句指的是「連…都/也…」這個結構。關於這個結構的本體研究較多，例如，崔希亮（1990）、曹逢甫（1994）、劉丹青、徐烈炯（1998）、洪波（2001），但在教學與學習方面的討論則相對地少。多數語法書和教材對

這個結構的說明是一種用於強調時的表達方式（呂叔湘 1980）。鄧守信（2009:115）則指出這個結構的語義可對應於英語的 *even*，較簡單；但在結構上較困難，因為「連」後面的賓語較多元，除了名詞（詞組）外，還包括狀語和句子。既然賓語多元，本文以「成分」取代「賓語」，不討論「連」的詞性問題。

肖奚強（2009）在這個結構的分析上，包括了「連」省略的用法，例如，「一句話也說不出來」。而我們在這小節中，僅針對有「連」的結構來討論，因此該書中關於省略用法的數據，我們都略過不提。該書 L1 與 L2 使用這個結構的比例分別是 0.019%（1,013 次/520 萬字）、0.015%（160 次/110 萬字），差距不大。在平衡語料庫與 TOCFL 語料庫中的使用率則是 0.027%和 0.030%，如表 7 所示，顯示差距也是不大的。

表 7 連字句在 TOCFL 學習者語料庫中的分佈

級別	使用次數	使用率 (%)	偏誤率%
A2	18	0.015	38.89
B1	138	0.031	23.91
B2	114	0.037	7.89
C	94	0.036	7.45

與把字句、被字句比較，不一樣的是，在 B1 階段，學習者連字句的使用率就已經與 L1 相當了，雖然偏誤率仍高，顯見這個結構對學習者而言，實用性較高，在語用上也因為沒有其他形式可以替代，學習者如果要表達這個語義一定得使用這個結構。

3.3.2 偏誤率

肖奚強（2009）所統計的平均偏誤率約 16%（數據歸納自該書 88 頁表 6.2），TOCFL 統計的平均偏誤率約 19%，從表 7 可以看出 A2、B1 學習者的偏誤率相當高，但 B2 以上則降到 10%以下，顯示 B2 學習者已經可以掌握這個結構。只是 B2 和 C 之間的差異不大，A2、B1 和 B2 之間的差異則是具有顯著性。這個結構可以做為[A2]、[B1]和[B2、C]這三群學習者的區辨特徵。

3.3.3 句式使用情形

在 A2 語料庫 18 筆用例中，15 個用例都是否定句式，例如：

*中文連個字我也不會讀 （正：連一個中文字我也不會讀）

18 筆中，其中 13 筆都是強調賓語、只有 2 筆強調主語，3 筆是謂語（連+動詞組），如「連吃東西，她都不能自己吃」。偏誤發生的情況多是因為語序問題，例如，「*我現在工作很忙碌，連一點也沒有休息的時間」。

在 B1 語料庫 138 筆用例中，強調賓語的仍然佔多數，出現不少「連+一+量詞+(名詞)+都/也+否定詞」結構，如「人民連一塊錢都不想用」；「連」字後面接的成分比較複雜了，如「連半夜去商店買隱形眼鏡或去藥房買藥都…」；出現在複句遞進結構中（連…都/也…，（更）別說/不用說…），如「連上廁所都很困難，更別說出去，我怎麼去上課呢？」，不過只有 1 筆。

在 B2 語料庫 114 筆用例中，首次出現 3 筆強調狀語的情況，如「因為連晚上我們也不能好好地休息」；能掌握使用這個結構「連…都/也…，（更）別說/不用說…」，如「我連找到工作都很難了，更別說找到好工作」；與「甚至」連用，如「甚至連好好地吃頓飯也不行」。整體而言，強調主語與謂語的情況都增加了，不過仍以賓語為主。基本上，這階段已經能掌握連字句的用法。

從上面的描述，可以大致瞭解每個階段學習者的習得情況，B2 以上學習者已經能自如地運用連字句，只是在所有語料中並沒有發現如 L1 使用的「連+V+都/也+不/沒+V」結構，如「連看都不看」、「連問都不問」（平衡語料庫）。推測是因為教材中，並沒有針對這種結構做特別的介紹，或是寫作主題的關係，沒有誘發使用的情境。

4. 結語與未來工作

除了上述三個句式外，我們另外也觀察了常用的「除了句、讓字句、爲了句」等等，相關資訊統整於表 8。可以發現在觀察 L2 與 L1 的使用率時，應該有針對性的來對比才是客觀的，例如，從表 8 中雖然得知 L2 被字句的平均使用率低於 L1，但 B2 階段的使用率其實已經趨近 L1，所以如果單純從平均數據顯示的過度使用或不足使用來論定 L2 是泛用或迴避使用都不恰當。

另外，我們也發現純粹從使用率來談，L2 的表現是否趨近 L1，也是不足的。從表 8 可以知道，像讓字句、爲了句等，是學習者超用的句式，在 A2 到 B1 階段的使用率就已經趨近 L1，但一直要到 B2 才能習得（「爲了」的語序）；連字句也是在 B1 階段使用率即趨近 L1，但要到 B2 才能習得。讓字句的習得階段尚未完成偏誤分析，表 8 裡以 N/A 標記。

可見在分析語法特徵時，這兩個數據都應參酌，從偏誤率可以大致瞭解 L2 習得情況，推測這個語法點的難度；從使用率則可以看到這個語法點或句式的實用性。例如，把字句的量化數據告訴我們，這個句式難度是相當高的，學習者要到 C 等級才能使用得好；另一方面頻率的數據則告訴我們把字句是相當實用的句式，學習者在 B1 到 B2 這個階段使用的頻率就已經趨近 L1，並不因為它的難度而阻礙了它的使用。連字句也是一樣的表現。

表 8 學習者常用特殊句式使用率（由高至低）

	讓字句	把字句	爲了句	被字句	除了句	連字句
L2 使用率	0.271	0.138	0.118	0.096	0.046	0.030
L1 使用率	0.135	0.144	0.062	0.148	0.049	0.027
L2 趨近 L1 階段	A2-B1	B1-B2	A2-B1	B2	B1	B1
習得階段	N/A	C	B2	B2	N/A	B2

從上面討論，我們可以更清楚使用率和偏誤率是兩個不同概念，在討論學習者交際所需或教學排序時，語法點或句式使用率的訊息更加有參考價值；在評量語法點難易度時，偏誤率及學習者習得該語法點的階段則可以做爲參考指標。

這個研究只是一個開始，雖然可以看到初步嘗試是可行的，這些常用的句式的確具有區辨不同能力學習者的功能，不過，我們應該要清楚界定這些句式下的次類結構，才能細緻地分析學習者表現，只是現階段還沒辦法做到。這個部分鄧守信（2009）在教學語法中曾指出，語言學家不需要考慮教學「累進式」的問題，但是對華語教學界而言，語法次結構的建立是當務之急。我們期望累積更多的研究成果，未來能有助於華語教學大綱的制定和教學材料的編寫，爲教材、教學目標、語法結構的編排等提供科學的根據；亦可反饋給考試單位，以利進一步修訂考試詞彙大綱、句法大綱。

引用文獻

- 中央研究院資訊科學研究所、語言學研究所、中央研究院計算中心網頁，中央研究院平衡語料庫。網址：
<http://app.sinica.edu.tw/cgi-bin/kiwi/mkiwi/kiwi.sh>
- 呂叔湘，1980，《現代漢語八百詞》，北京：商務印書館。
- 李婕、陳晨，2012，〈基於 HSK 動態作文語料庫的被字句習得考察〉，《國際漢語學報》3.1:109-120。
- 肖奚強 主編，2009，《外國學生漢語句式學習難度及分級排序研究》，北京：高等教育出版社。
- 洪波，2001，〈連字句續貂〉，《語言教學與研究》2:55-63。
- 桂孟秋，2011，〈面向留學生的被字句語用教學—基於 HSK 動態作文語料庫的研究〉，《江漢大學學報（人文科學版）》30.1:65-68。
- 崔希亮，1990，〈試論關聯形式「連…也/都」的多重語言信息〉，《世界漢語教學》3:139-144。
- 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，2013，〈TOCFL 作文語料庫的建置與應用〉，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》141-152，北京：北京語言大學出版社。
- 張燕吟，2003，〈準確率標準和初現率標準略談〉，《世界漢語教學》65.3:52-61。
- 張寶林，2010，〈迴避與泛化—基於 HSK 動態作文語料庫的把字句習得考察〉，《世界漢語教學》24.2:263-278。
- 張寶林、崔希亮、任傑，2004，〈關於「HSK 動態作文語料庫」的建設構想〉。2010年3月15日，網址：
http://china-language.gov.cn/54/2007_6_20/1_54_1356_0_1182321111359.html
- 曹逢甫，1994，〈再論話題和「連……都/也」結構〉，戴浩一、薛鳳生 主編《功能主義與漢語語法》95-116，北京：北京語言學院出版社。
- 劉丹青、徐烈炯，1998，〈焦點與背景、話題及漢語「連字句」〉，《中國語文》4：243-252。
- 蔡雅薰，2009，《華語文教材分級研制原理之建構》，臺北：正中書局。
- 鄧守信，2009，《對外漢語教學語法》（修訂二版），臺北：文鶴出版社。
- 謝妙玲、林雅惠，2008，〈「歐洲語言學習、教學、評量共同參考架構」在中

- 文教學方面的應用：國立台灣師範大學國語教學中心的經驗〉。發表於 2008 年全美中文教師學會年會（2008.11）。美國：佛羅里達州 Orlando。
- 藍珮君，2007，〈基礎華語文能力測驗與歐洲共同架構的對應關係〉，發表於第三屆華文教學國際論壇（2007.12）。臺北：國立臺灣師範大學。
- American Council on the Teaching of Foreign Languages [ACTFL]. 1987. ACTFL Chinese Proficiency Guidelines. *Foreign Language Annals* 20.5:471-487.
- Brown, Roger. 1973. *A first language: The early stages*. Cambridge, Mass.: Harvard University Press.
- Capel, Annette. 2010. A1-B2 vocabulary: insights and issues arising from the English Profile Wordlists project. *English Profile Journal* 1 e3 doi: 10.1017/S2041536210000048
- Chen, Ching-Yu, Shu-Fen Tseng, Chu-Ren Huang and Keh-Jiann Chen. 1993. Some distributional properties on Mandarin Chinese- A study based on the Academia Sinica Corpus. *Proceedings of Pacific Asia Conference of Formal and Computational Linguistics I*, 81-95.
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. 2005. *Guide for the production of RLD (Version 2)*. Strasbourg: Language Policy Division. Available online, November 2005. http://www.coe.int/t/dg4/linguistic/Source/DNR_Guide_EN.pdf
- Granger, Sylviane (ed.). 1998. *Learner English on computer*. London and New York: Longman.
- Hawkins, John A. and Paula Buttery. 2010. Criterial features in learner corpora: Theory and illustrations. *English Profile Journal* 1.1:1-23. Cambridge University Press. doi:10.1017/S2041536210000103
- Pienemann, Manfred. 1998. *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
- Salamoura, Angeliki and Nick Saville. 2009. Criterial features of English across the CEFR levels: evidence from the English Profile Programme, Cambridge ESOL. *Research Notes* 37:34-40.

The usage of some sentence patterns by L2 Chinese learners: A corpus-based study

Li-ping Chang

National Taiwan Normal University

Abstract

This study aims to explore the salient linguistic features of Chinese syntactic structures across the CEFR levels. It takes Ba sentences, Bei sentences, and Lian sentences as examples to examine their usage in terms of occurrence frequency, errors, and sub-structures produced by L2 Chinese learners. The research is based on corpus linguistics method, by comparing the sentences in a learner (L2) corpus and a native-speaker (L1) corpus. The L2 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 L1 corpus is the Academia Sinica balanced corpus. According to the statistical significance criterion proposed by Hawkins and Buttery (2010), these specific structures are used as the criterial features to classify different proficiency levels.

Keywords: Mandarin Chinese, sentence pattern, criterial feature, learner corpus, error analysis