

07

CHAPTER

TOCFL學習者語料庫的 偏誤標記

◎張莉萍
臺灣師範大學

7.1 研究背景

對學習者語料庫而言，載有偏誤訊息是非常重要的特徵，目前學習者語料庫偏誤標記的方式，有兩大系統（Tono 2003），一是語言學分類系統（linguistic category classification, LCC），一是目標語差異系統（target modification taxonomy, TMT）。

早期TOCFL學習者語料庫為了便於研究者分析（張莉萍 2013），也做了簡要的偏誤標記工作，例如：

- (1) 老師一看到他就很生氣了[R]。不僅是這樣，他的功課還沒做完。
- (2) 很美[M]風景。

(1) 中的[R] (redundancy) 指的是冗贅，這裡顯示學習者多用了「了」；(2) 的[M] (missing) 指的是遺漏，這裡學習者少用了「的」。這種標記方式採用的就是TMT，可以知道學習者語言與目標語言差異的地方，但沒有進一步資訊可以分析學習者多用了什麼或少用了什麼。一般而言，愈精細的分類，愈有助於二語習得研究或偏誤分析，然而，分類愈細，也有不利之處，例如，標記人員對標記的掌握能力，將影響標記工作的一致性。但從上述例句的做法也可以看出，過度簡化的標記，一旦要實際做大量觀察或量化分析時，並無法達到研究者的需求，也使研究者必須針對不同的研究主題，重新檢視文本，甚至重新標記新的訊息，如此不斷地重新檢視、標記，著實耗費了大量精神與時間。因此，經過一年的實作，我們重新思考如何讓語料庫加註偏誤標記的這項工作，能做到全面性與有效性。

所謂全面性，是希望這套偏誤標記可以涵蓋學習者偏誤的各種表現；所謂有效性，有兩個層面：一是指透過偏誤標記，使用者可以很有效地分析語料；二是指標記結果的一致性和精確性，唯有如此，偏誤標記才是可靠、有效的。要達到這兩個目標並不容易，與偏誤標記的設計、標記工作的步驟、流程皆息息相關。

本章第7.2節將先進行偏誤分類的相關文獻探討；第7.3節則說明TOCFL語料庫偏誤標記設計的理念與方法；第7.4節說明進行偏誤標記的步驟流程；最後做一總結。

7.2 偏誤分類的文獻探討

關於偏誤分類的相關文獻，可分三部分來談：一是從理論層面探討偏誤分類的問題；二是從實務層面來看語料庫偏誤標記的分類方法；三是檢視目前華語學習者語料庫的偏誤標記工作。

7.2.1 偏誤分類理論

討論學習者偏誤類型的文獻相當多，較早有系統性討論的當屬 Dulay, Burt, and Krashen (1982:Ch. 7)，他們從語言學分類、表面策略 (surface strategy)、比較分析及交際效果四個方面來討論。

語言學分類即是依據語言學概念來為偏誤做描述性的分類，例如，偏誤可以分成構詞層面或句法層面。句法中可再區分為定詞偏誤、名物化偏誤等類型分項敘述。定詞下可再依遺漏、錯用等項目來敘述。

表面策略分類則是從學習者句子的表面結構和目標語差異的狀況分成四類，分別是遺漏 (omission)、誤加 (addition)、形式錯誤 (misformation) 和錯序問題 (misordering)。

比較分類法是基於L2偏誤結構和某些其他類型結構的比較，例如，和目標語習得結構或學習者母語結構的比較，通常可分為：(1) 發展性偏誤 (developmental errors)，這類偏誤類似目標語兒童習得語言的階段性錯誤；(2) 語際性偏誤 (interlingual errors)，這類偏誤起因於學習者母語結構遷移；(3) 歧義性偏誤 (ambiguous errors)，指的是偏誤來源可能是前述的發展性偏誤，也可能是語際性偏誤；(4) 其他類，也就是不能歸因於前述偏誤來源者。

交際效果的分類法不同於前三項分類法，所關注的是會產生溝通問題的偏誤和不會產生溝通問題的偏誤。Burt and Kiparsky (1972) 將其分為兩大類：一是整體性偏誤 (global errors)，偏誤主要類型包括句子主要成分語序的偏誤和句連接詞的錯用、遺漏和錯序，例如，「*English language use many people.」、「*He started to go to school since he studied very hard.」；二是小範圍偏誤 (local errors)，是句中單一成分的偏誤，例如，「*Why we like each other?」，此類偏誤不影響溝通。

Dulay, Burt, and Krashen (1982) 指出使用表面策略分類法來分析

偏誤，讓研究者較能系統性地關注學習者的偏誤，因為從表面結構的差異，可以比較清楚看到學習者在重建一個新語言時認知的過程。我們也可以看到後來研究者多引用這類分析法，例如，被廣為引用的James（1998），作者只是將Dulay等人的surface strategy這類改稱為target modification taxonomy，因為他認為這類基本上就是表達中介語和目標語之間的差異，他在四類之外，另加了「混合」（blends）類，定義為不只一個可能的目標語形式，如「*according to Erica's opinion」可能修正為「according to Erica」，也可以修正為「in Erica's opinion」（James 1998:111），因此形成五小類：遺漏、誤加、錯用（misselection）（原Dulay等人的misformation）、錯序和混合。從上述文獻顯示，Dulay等人在1980年代已為偏誤分類做出了全面性的描述，經過James進一步地修正，這種分類法以及術語已為多數學者或語料庫建置者沿用至今（詳見7.2.2節）。

不過，對於James加了「混合」類，筆者認為並不合適。我們實際去看學習者的真實語言就可發現，學習者的很多產出在目標語中的表達方式可能不只有一種（肖奚強 2011；張莉萍 2013；Barlow 2005），若要標記者去推測這個偏誤可能對應到目標語兩種以上的表達方式，既不精確，也無益於量化與分析，反而容易淪於標記時的垃圾桶。

7.2.2 語料庫偏誤標記的類型

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1,000萬詞規模，作為商業用途。Turton and Heaton（1997）也利用這個語料庫編著了學習者常見偏誤辭典。另外，有劍橋學習者語料庫（Cambridge Learner Corpus, CLC），約4,000萬詞，語料來源是劍橋考試中心所舉辦的寫作考試文本，考生超過16萬，來自不同國家的英語學習者，其中對半數語料進行了偏誤標記（Nicholl 2003）。而目前國際上最有計畫，廣泛地搜集語

料，也最為人所知的是國際英語學習者語料庫（International Corpus of Learner English, ICLE），主要負責人是比利時的Granger教授，語料是由每篇大約700字的進階級作文所組成（Granger et al. 2009）。這些作文基本上都是從不同國家的課堂上取得，目前有17個合作夥伴國加入，也為語料做了偏誤標記（Barlow 2005:338）。關於其他學習者語料庫的介紹可參閱本書第一章1.3.2及1.5.3小節。

對語料庫偏誤標記的工作，Diaz-Negrillo and Fernandez-Dominguez（2006:87）回顧了1990~2005年間學習者偏誤標記系統的發展，歸納結果顯示LCC類標記是每個語料庫系統都標示的，如比利時魯汶大學（Katholieke Universiteit Leuven）所開發的FreeText系統有11個主要的詞類、54個次分類、12個標點符號類，共計100多個標記符號；日籍學習者英語語料庫（NICT JLE）（Izumi, Uchimoto and Ishahara 2005）有11個詞類，共46個標記；CLC有9個詞類，總共有88個不同標記（Nicholls 2003）。以下是CLC偏誤標記的例子：

(3) *He died<#MP>|.We </#MP> buried him the next day.* (Nicholls 2003:574)

<#MP>是omission of punctuation的縮寫，表示標點遺漏，並將標點「.」補上。這個標記可以看出有兩個概念：一是遺漏（M），屬於TMT概念；一是標點（P），屬於LCC概念。但不是每一個偏誤標記的系統都完整呈現這兩個概念，例如，NICT JLE幾乎都是LCC分類系統，在無法歸入詞類的偏誤時，才設置了其他類，如錯序，而這樣的標記僅有5個。詳見Izumi, Uchimoto, and Ishahara（2005）的表1。偏誤標記的例子如下：

(4) *I belong to two baseball <n_numcrr= "teams"> team </n_num>.*
(Izumi, Uchimoto and Ishahara 2005:75)

例中n表示名詞類錯誤，num表示數量類偏誤，在引號內的是更正後的訊息。從這個例子中，可以看出只有LCC的分類概念。

從現今國際上大型學習者語料庫的偏誤標記工作可以得知，偏誤標記基本上採用了自Dulay, Burt, and Krashen (1982)、James (1998)以來的分類架構，也就是TMT和LCC兩大概念，TMT的分類沒有太大彈性，就是遺漏、冗贅、錯序、詞語選用錯誤四大類；LCC則因各系統對語言學分類的粗細，數量差異相當大，從幾個到100多個標記都有，當然標記的粗細，視各語料庫的研究性質而有不同。

從各語料庫的標記系統，我們也可以看到多數的標記集 (tagset) 都有階層性的概念。例如上述 (4) 例中，n是主要類別，屬於數量的num則是它下一層次的次類。ICLE的標記集也是如此設計，主類有七大類，其中G代表語法類，GV則是語法類中的次類「動詞」；GVT代表更低階層的時式偏誤類 (Dagneaux, Dennes, and Granger 1998:166)。在第7.3節，可以看到TOCFL的標記集也有階層性，目的是為了讓標記人員在判斷時能更清楚脈絡，增加標記時的效率與一致性；另一方面，也讓標記集更有彈性，也就是主類不會變動，但次類可以因為研究的需求而具有擴充性。

7.2.3 華語學習者語料庫偏誤標記情況

上述所描述的學習者語料庫都是以英語為目標語，關於漢語帶有偏誤標記的學習者語料庫則以中國大陸的「HSK動態作文語料庫」為代表 (張寶林、崔希亮、任傑 2004)。這個語料庫的內容是母語非漢語的外國人參加高等漢語水平考試 (HSK高等) 所寫的作文，於2003年開始建置，2006年年底試行開放網路版本1.0，至2009年正式公開1.1版本。搜集了1992~2006年部分外國考生的作文答卷，共計11,569篇，424萬字。從字、詞、句、篇、標點符號等角度分類，約50種錯誤標記。

該語料庫依據偏誤所在位置的語言成分分為字、詞、句、篇章

不同層次，層次分明，但在實際操作時難免受到考驗，我們從下列的(5)～(7)的例子來進一步說明。

- (5) ……有的農民{CQ在}不使用化肥和農藥的情〔B精〕況下
[BD,]養農作物……。
- (6) ……現代社會應當認「保護自己，尊重其他人」為口號
{CJgd}……。
- (7) ……我認為這種現象……在很多人身上{CJ-zy都}會有的。

(5)～(7)是摘錄自HSK動態作文語料庫中的「語料標注與代碼說明」中的例子 (<http://202.112.195.192:8060/hsk/help2.asp>)。例句(5)的標記，說明文件中標示為缺詞(CQ)，意思是考生遺漏了「在」字，判斷這個偏誤是在「詞」的層次上。而例句(6)則是「句」層次上的偏誤，CJgd代表的是固定格式上的錯誤。例句(7)也是在「句」層次上的偏誤，-zy代表的是狀語成分殘缺。

事實上，就HSK的定義，如果我們把「在……下」當成一個結構來處理，那麼例句(5)的偏誤也可以視為是在「句」的層次上，而非「詞」的層次。而例句(6)，考生所犯的偏誤可以解釋為誤將「認為」這個詞當做可以拆的詞語結構，也就是離合詞，那麼應該屬於「詞」的層次，而非「句」的層次。另外，HSK在句成分殘缺或多餘(CJ-)的標記之下又以主語、謂語、狀語等功能分為八小類，將狀語成分殘缺(zy)視為「句」層次，而非詞層次的缺詞問題(CQ)，也值得商榷，如例句(7)，只因為「都」可以做狀語，而歸類為句層次。但這樣的處理沒有太大的意義，漢語可以充當狀語的詞語太多了，標記狀語缺失沒有太大意義，偏誤的關鍵還是在於學習者不能掌握「都」的使用，該用而沒用。

從上述這些例子可以看到，HSK語料庫偏誤標記基本上還是基於語言學分類的考量較多，例如，{CJ-zy}中CJ是句層次，zy是狀語，都

是LCC概念。只是LCC的概念多呈現在句的層次，如上所述，有主語、狀語、賓語、定語等，另外也有句式的標記，如把字句、被字句、存現句等，但沒有詞類的細緻標記。部分標記則涵蓋了TMT概念，例如，{CQ}指的是缺詞（即遺漏）。雖然標記工作到達篇章範圍，但實際上關於篇章層次的標記只有一個，即{CP……P}，意義並不大。因此，在標記的架構上劃分詞、句、篇章等層次是否有必要，值得商榷。

不過，這個語料庫對於大部分的偏誤語料都做了更正，也就是可以同時找到正確的表達形式，有利於精確地檢索。該檢索系統介面可以參考本書第一章圖1.10。

7.3 TOCFL語料庫的偏誤標記理念

在本章7.1節提到TOCFL語料庫在早期做過簡化的偏誤標記，經過一年的實作，體認到這項工作的複雜與辛苦。雖然我們一開始也訓練了人員對偏誤標記的認識以及標記的原則，但囿於人力與時效，並未採行兩人同時標注一份檔案的方式，做到互相校對的程序。之後檢視，發現最大的問題即是標記不一致，以及太過簡化的標記使得量化結果過於粗糙，這讓我們重新思考該如何讓語料庫加註偏誤標記的作業能做到全面性與有效性。以下將從這個帶有偏誤標記的語料庫所針對的使用對象、偏誤標記的內涵、標記的範圍等面向來說明這套偏誤標記集。

7.3.1 標記的內涵與範圍

對於學習者語料庫帶有偏誤標記的必要性，見仁見智，端看使用者的目的。對一般使用者而言，原始的文本與簡單的檢索功能可能就可以符合需求，可大致觀察學習者產出的語言是什麼樣子；對於進階的研究者而言，大量的語料若未經整理、分類，在使用上仍是非常不便，也很難看到全貌，例如，學習者缺漏的語言成分無法經由詞語檢索得到。因

此，當決定進行這項偏誤標記作業時，這個語料庫設定的對象就是進階研究者。然而，這又牽涉到另一項議題，畢竟每個人研究的興趣不同，要做到理想的全面性標記談何容易？例如，對於語體、語用或篇章研究者而言，現階段的標記內容仍不足以提供此方面的訊息。

那麼怎麼讓此標記系統有擴充的可能性，能顧及研究者的個別需求呢？我們採取階層式的分類方式，不同於HSK語料庫的標記方式，第一個階層是TMT分類方法。也就是，先從學習者的語言和目標語的差異分成四類：遺漏、冗贅、選用錯誤和語序錯誤，分別以M、R、S、W這四個符號為代表。這個階層的標記是固定不變的。第二個階層則是LCC分類法，也就是每一個偏誤給定了主類後，根據語言學概念給予次類。此階層的標記是有擴充性的，現階段依據一年來的實作經驗，就已觀察到的現象，規範了36個次類，這36個次類分布在詞、語、句型和其他這四個範圍內，如下所示：

1. 「詞」的範圍（16類）¹，如動作動詞（v）、能願動詞（aux）、狀態動詞（vs）、名詞（n）、代名詞（pron）、連接詞（conj）、介詞（p）、數詞（num）、指示詞／定詞（det）、量詞（cl）、句末助詞（sp）、時態助詞（asp）、副詞（adv）、結構助詞（de）、疑問詞（que）、複數詞綴（plural）。

2. 「語」的範圍（11類），如主語（sub）、賓語（obj）、名詞短語（np）、動詞短語（vp）、介詞短語（pp）、修飾語（mod）、時間詞語（time）、地方詞語（loc）、及物性（tran）、[數／定＋量]結構（dm）、離合結構（vo）。

3. 句型範圍（7類），如複雜名詞子句（rel）、特殊句型（把ba、被bei、讓rang、是shi、有you）、其他句式（pattern）。

¹ 我們的標記有一小類是要標示學習者使用詞綴「們」的問題，由於只有此類屬於詞綴，為簡化呈現，將它歸入「詞」的範圍。

4. 混合範圍（2類），如形式錯誤（form）、表達方式錯誤（sentence）。

之所以細分為36類，是為了未來能提供細緻的分析與統計。在文獻探討中我們亦提及，有的學習者語料庫的標記符號多達一百多個，端看研究者的興趣或範圍。基本上，我們的標記亦有一百多個，因為假設每個主類之後都跟著36個次類，那麼就有144個不同標記。但對標記者或使用者而言，只要在第一階段區分四大類，再根據詞性（16類）或語類、句型等給予次類，把36個類別掌握好，其實不需要記憶144個類別。這正是階層式分類的好處，除了便於記憶，也具有系統性、擴充性。例句（8）～（11）的例子即是展示不同主類下的同個次類。

（8）公共汽車沒有空的椅子，所以兩個人都得站[Masp]。

→公共汽車沒有空的椅子，所以兩個人都得站著。

（9）我朋友也找工作找了很久才找到了[Rasp]。

→我朋友也找工作找了很久才找到。

（10）他爬山爬的[Sasp]半天。

→他爬山爬了半天。

（11）我放假的時候去歐洲了[Wasp]。

→我放假的時候去了歐洲。

這四個句子都是關於時態助詞（asp）的偏誤表現，分別是少用、多用、選用錯誤和語序錯誤的問題。

7.3.2 次類的設計理念與標記原則

針對LCC分類（即次類）的精緻性，視研究的需求而定。在這個系統中，我們不標記學習者用字的問題，因為這個語料庫的來源是電子文本，也就是學習者都是經由鍵盤輸入中文，漢字問題與研究手寫漢字偏誤的類型不同，多是「選字」錯誤所引起，關於漢字的問題，為

了不影響自動斷詞的準確率，在建置語料庫前即做了特殊的前置處理（張莉萍，2013:146）。在偏誤標記時，略過字的偏誤，標記最小的語言單位是詞綴，例如，「們」[plural]。大部分的偏誤都是在詞的範圍，這時除了判定四個主類外，需要進一步標記出該詞的詞類，如上述例句（8）～（11）的例子。至於詞類究竟要給到多細？我們除採用鄧守信（2009）八大類（即名詞、動詞、副詞、助詞、連詞、介詞、限定詞、量詞）的概念之外，另特別標記出時態助詞、結構助詞、句末助詞、代名詞、疑問詞等，主要是這些虛詞或帶有特別功能的詞語，是學習者的難點，也是研究者熱衷觀察分析的語言點。

實際上，學習者所產出的語言不是只有選用詞語或使用詞語正確與否這類單純的問題。例如，當學習者寫出「我慶祝會不能參加」這樣表面正確的句子時，實際上可能是錯誤的表達。因為從上下文可以知道，在行文中，「慶祝會」是第一次出現，之前並沒出現過「慶祝會」這個訊息，所以，正確的表達應該是「我不能參加慶祝會」。此時在語料庫中的標記方式就如例句（12）所示，次類的標記為[obj]而非[n]。這時如果因「慶祝會」是名詞而標記[Wn]則失去偏誤分析的訊息。因為這不是個別名詞的偏誤問題，而是學習者無法掌握賓語的位置。這也是為何在詞類之外，標記也涵蓋了「語或詞組」的範圍。

（12）我慶祝會[Wobj]不能參加。

→我不能參加慶祝會。

（13）我們到現在都是以寫信為連絡[Mnp]。

→我們到現在都是以寫信為連絡的方式。

例句（13）也是詞組範圍的一個標記例子，學習者缺失了「的方式」，這時缺失的單位大於名詞，則選擇次類名詞組[np]的標記。在詞組範圍內，除了主語、賓語、介詞組、名詞組等標記外，還有一個比較不常見的標記，即修飾語[mod]，指的主要是[X+]的結構，這也是二語

習得研究者所關心的主題，在系統中特別標示出學習者在這一類的偏誤。如例句（14）～（16）所示，如果僅標示NP、VP，便無法顯示出詞組的功能，而失去偏誤訊息。研究者可以檢索出這類[mod]，較省力地針對修飾語的多用、少用、錯序等進行觀察分析。

（14）現在我知道甚麼樣的[Wmod]買東西是值得的。

→現在我知道買甚麼樣的東西是值得的。

（15）因為你要到我的城市來看看[Mmod]生活怎麼樣。

→因為你要到我的城市來看看這裡的生活怎麼樣。

（16）我希望你改掉這兩個你的[Rmod]生活習慣。

→我希望你改掉這兩個生活習慣。

在華語教學中，特殊句型也是教學的重點項目。在此系統中，我們針對把字句、被字句、讓字句、是字句、有字句這五種句型做了標記。在標記原則中，特別說明偏誤牽涉到這些句式時，不論是多用、少用、錯序或錯用都要優先標示ba、bei、rang、shi、you，而非詞類標記。另外，在句的範圍內，還有一個現象也是語言學者所熱衷探討的，那就是複雜名詞子句或稱關係子句，我們標記為rel，如例句（17）～（19）所示。這類偏誤若未透過人工標記出來，很難進一步進行量化、質化分析。

（17）我覺得餐廳在你家旁邊[Srel]好像好吃。

→我覺得在你家旁邊的（那家）餐廳好像很好吃。

（18）希望我們的計畫要見面[Srel]沒有甚麼問題。

→希望我們要見面的計畫沒有甚麼問題。

（19）我覺得野餐是最好的辦法認識新朋友[Srel]。

→我覺得野餐是認識新朋友的最好辦法。

上述所討論的標記大致涵蓋了詞、詞組、句子的範圍，還有兩個次

類——[form]、[sentence]很難納入這些範圍內，姑且稱這兩類為其他類或混合類。關於形式（form）的偏誤，歸納有以下三種情況：

1. 選用的詞沒錯，只是形式錯誤（例如重疊式、嘗試貌），如例句（20）、（21）。
2. 可以理解語義，只是表達形式錯誤，如例句（22）～（24）。
3. 遺漏了固定格式中的某部分，如例句（25）。

（20）他在路上走著想想[Sform]他的功課。

→他在路上走著想他的功課。

（21）誰都不知明天也[Radv]自己的工作或公司還會[Sform]存在。

→誰都不知明天自己的工作或公司還會不會存在。

（22）幸運[Mform]我住的地方有越南餐廳。

→幸運的是我住的地方有越南餐廳。

（23）李文平敬於[Sform]。

→李文平敬上。

（24）我覺得不舒服得不得了[Sform]。

→我覺得非常不舒服。

（25）可是對他們[Mform]是開心的一天吧！

→可是對他們來說是開心的一天吧！

1類的標記方法可以有別於學習者選用詞的標記（例如，Sv或Sn），因為這類偏誤的重點並非選用了錯誤的詞，而是詞的構形問題。2類的用意則是標記出學習者使用錯誤形式，而這個錯誤形式根本不是一個詞，如例句（23）的「敬於」。3類則是有別於單純標記缺失了名詞組或動詞組，因為這裡所缺失的是固定格式中的某個成分。無庸置疑地，[form]這一類搜尋出來的偏誤語料會較紛雜，但因為初始標記時，無法預測各小類的數量，又不想讓標記符號擴張太大，造成標記者負擔，於是，採用單一標記來記錄這幾小類。

[sentence]這個標記是指學習者的語言需要做整體性的修改，前提是標記者（母語者）能理解意思。如果透過上下文，標記者仍不得其門而入，只能放棄不標。透過這個標記，研究者至少可以進一步觀察分析，學習者採用何種策略來產出想要表達的句義。如例句（26）～（27）。

（26）合格了交換學生[Ssentence]以後

→符合了交換學生的資格以後。

（27）她是最好的家教一樣的感覺[Ssentence]

→感覺她是最好的家教一樣。

在這一小節，我們簡要地敘述了標記系統的架構以及設計理念，還有許多細節無法在此一一描述，有興趣的讀者請詳見附錄關於此系統的說明手冊，手冊中詳列了標記的類別、意涵、例句以及標記的大小原則、標記選用的優先順序等。這個手冊除了提供給系統使用者參考外，更是培訓標記人員時必要的文件。

7.4 偏誤標記的步驟與結果統計

在進行偏誤標記工作之前，研究者必須招募人員、訓練人員、製作偏誤標記說明手冊。人員的素質、訓練的過程與手冊的說明決定了未來工作的成敗。在人員部分，理想的人選需要具有語言學知識背景以及華語教學經驗，因為標記者既要有漢語語言學概念，也要對學習者語言有初步的體認。

至於人員的訓練，如何確定經過研究人員的講解，在實際標記時，能達到一定的準確性，是我們所關心的問題，這關係著標記結果的有效性。然而，偏誤標記這項工作，既然是人工進行，難免帶有主觀性，雖然製作操作手冊就是為了把主觀性降至最低，但不同能力學習者的偏誤

表現仍遠超過手冊所能規範。理想的步驟應是兩人共同標記一份文檔。但是偏誤標記是一項費時、費工的作業，在有限時間與人力下，這個理想幾乎是遙不可及的。於是我們採取了折衷做法，讓人員在研讀手冊、接受訓練之後，每個人標記同樣的50篇作文，而這50篇作文，研究者事先做好了標準版本，程式設計人員為我們設計了一個工具，可以標示出每個人和標準版本不一致處，這不一致包括對偏誤的認定以及對偏誤標記的指定，也就是可能有人認為是偏誤，給了標記符號，但另一個人不認為是偏誤，沒給標記；也可能是兩個人都認定是偏誤，但給的標記類別不一樣。標記不一致的結果輸出後提供訓練者檢視，訓練者再針對新進標記者對標記的認知模糊處，進行再教育。這樣的訓練期大概需要兩週的時間，這兩週需要訓練者、標記者與研究者之間密切地來回討論，而從這個過程中亦可顯現說明手冊在實際操作時的不足處，研究者必須不厭其煩地檢視修訂其內容。

標記人員經過訓練之後，正式標記時，使用的是單機版的標記編修工具，是由程式設計人員為這項工作量身訂做的，主要目的是為了搜集編修後語料的便利性，以及減少人員因為打字輸入標記的意外錯誤。

圖7.1即是編修工具的視窗，最左側是選擇編修檔案的功能。選擇之後會出現兩個同樣的文檔，左邊文檔是插入偏誤標記所用，右邊文檔是更正學習者語言所用。舉例來說，在語料稱謂下的第一行顯示「這是我們要[Raux]第一次見面的[Rde]」，在右邊可以看到標記人員更正文本為「這是我們第一次見面」。編修完畢會同時儲存兩邊的檔案，日後可以使用插入標記的文本以及更正文本做相關的研究。



圖 7.1 偏誤標記工具電腦螢幕視窗

此工具的左上角還有一個Report功能，可以即時統計出目前經過偏誤標記的語料的所有偏誤類型與次數。例如，現階段我們總共完成了2,837篇作文的偏誤標記，點選Report功能鍵，會開啟出如圖7.2所示的統計條狀圖。

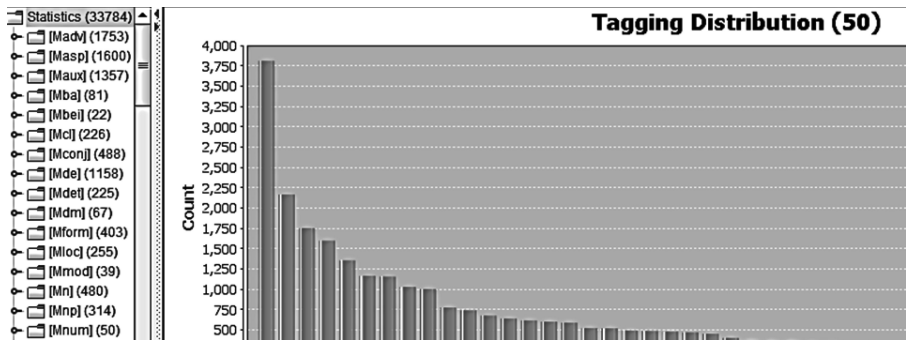


圖 7.2 偏誤標記工具統計畫面

從圖7.2可以看出目前共有33,784處偏誤標記。從長條圖可以看出最多次數的類型是[Sv]，也就是動作動詞選用錯誤，有3,811次。次高是名詞選用錯誤[Sn]，有2,166次。第三高則是副詞遺漏[Madv]，有1,753次。接下來4~10名依序是時態助詞遺漏（1,600）、能願動詞遺漏（1,357）、副詞選用錯誤（1,166）、結構助詞遺漏（1,158）、動作動詞遺漏（1,031）、時態助詞冗贅（1,003）、狀態動詞選用錯誤（776）。統計功能也可以僅針對36個次類做統計。現階段前10名次類如表7.1所示。

表 7.1 出現頻率最高的前10名偏誤次類

次類	v	adv	n	asp	de	Aux	conj	p	vs	form
次數	5445	4133	3001	2808	2334	2325	1418	1215	1154	1020

從表7.1可以看到三大詞類（名詞、動詞、副詞）居偏誤的前三名。這個不難理解，因為這三個詞類是漢語的前三大詞類，詞彙數量最多，犯錯的次數高也就不足為奇。不過，時態助詞和結構助詞分居第四名和第五名，則可以解釋為這兩類的確是學習者的難點，因為這兩類詞的數量很有限。另外，搭配圖7.2的統計圖來觀察，也可知動作動詞和名詞的偏誤呈現在四大主類中最多的的是選用錯誤這一類；副詞偏誤則是分布在遺漏類較多。這些都是概略的統計數據，未來研究者可從這些標記過的語料再縮小範圍去深入觀察分析。

7.5 結語

現階段TOCFL語料庫帶有偏誤標記的語料完成近100萬字，在2015年已將標注完成的所有語料放置在<http://tocfl.itc.ntnu.edu.tw>網站上，供大眾檢索使用，至今已累積一萬多瀏覽人次。由於標注工作相當耗時費

力，雖然距離全面性、有效性的目標還有一段距離，此項工作必須暫時告一個段落，未來將俟使用者的反饋，評估這項工作對於漢語作為第二語言的教與學的助益，再決定是否持續標注更多的語料。無論如何，期望這項工作的成果至少可以應用在語言教學上，提供二語教學者、學習者、辭典編輯者有用的參考資料，也可應用在中文計算機語言學研究上，作為訓練機器（電腦）找出偏誤的基礎材料（Lee et al. 2014）。

致謝

本研究感謝科技部計畫（MOST103-2410-H-003-043-MY2）、科技部「跨國頂尖研究中心計畫」（NSC103-2911-I-003-301）、教育部「邁向頂尖大學計畫」以及臺灣師範大學「華語文與科技研究中心」的支持。關於本研究所使用的工具，筆者特別感謝臺師大曾元顯研究員所帶領的團隊：李龍豪、李奎慶、廖柏舜等人開發比對軟體、偏誤編修軟體，提供研究團隊使用。關於偏誤標記工作，特別感謝鄭琬鈴、楊雅筑、楊志盛、劉美君的協助。本文初稿發表於第三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。

參考文獻

- 肖奚強，2011，〈漢語中介語研究論略〉，《語言文字應用》2：109-115。
- 張莉萍，2013，〈TOCFL作文語料庫的建置與應用〉，崔希亮、張寶林編：《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。
- 張寶林、崔希亮、任傑，2004，〈關於“HSK動態作文語料庫”的建設構想〉，《第三屆全國語言文字應用學術研討會論文集》544-554。
- 鄧守信，2009，《對外漢語教學語法》（修訂二版），臺北：文鶴出版社。
- Barlow, Michael. 2005. Computer-based analyses of learner language. In Ellis, Rod and Gary Barkhuizen (eds.) *Analysing Learner Language* 335-357. Oxford: Oxford University Press.
- Burt, Marina K. and Carol Kiparsky. 1972. *The Gooficon: A Repair Manual for English*. Rowley, MA: Newbury House.
- Dagneaux, Estelle, Sharon Denness and Sylviane Granger. 1998. Computer-aided error analysis. *System: An International Journal of Educational Technology and Applied Linguistics* 26.2: 163-174.
- Díaz-Negrillo, Ana and Jesús Fernández-Domínguez. 2006. Error tagging systems for learner corpora. *Resla* 19: 83-102.
- Dulay, Heidi C., Marina K. Burt and Stephen D. Krashen. 1982. *Language Two*. Oxford: Oxford University Press.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot. 2009. *The international corpus of learner English. Version 2. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Izumi, Emi, Kiyotaka Uchimoto and Hitoshi Ishahara. 2005. Error annotation for corpus of Japanese learner English. *Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora (LINC 2005)* 71-80.

- James, Carl. 1998. *Errors in Language Learning and Use: Exploring Error Analysis*. London: Addison Wesley Longman.
- Lee, Lung-Hao, Liang-Chih Yu, Kuei-Ching Lee, Yuen-Hsien Tseng, Li-Ping Chang and Hsin-His Chen. 2014. A sentence judgment system for grammatical error detection. *Proceedings of the 25th International Conference on Computational Linguistics (COLING'14)* 67-70. Dublin, Ireland, 23-29 August.
- Nicholls, Diane. 2003. The Cambridge learner corpus: Error coding and analysis for lexicography and ELT. *Proceedings of the Corpus Linguistics 2003 Conference* 572-581. Lancaster, United Kingdom, 28-31 March.
- Tono, Yukio. 2003. Learner corpora: Design, development and applications. *Proceedings of the Corpus Linguistics 2003 Conference* 800-809, Lancaster, United Kingdom, 28-31 March.
- Turton, Nigel D. and John Brian Heaton. 1997. *Longman Dictionary of Common Errors*. Harlow, UK: Longman.

附錄

TOCFL學習者語料庫偏誤標記說明手冊

標記分類法

主要參考Dulay, Burt, and Krashen (1982) 以及James (1998)。

四大主類

先從句子的表面結構和目標語的差異分成四類：

類型	與目標語的差異
遺漏錯誤 (M)	少用了語言成分
冗贅錯誤 (R)	多用了語言成分
選用錯誤 (S)	錯用了語言成分
語序錯誤 (W)	語言成分錯序

36次類

給定了主類後，根據語言學概念給予次類。這個次類又分為幾個層次，包括：

(1) 「詞」的層次 (16類)，如動作動詞 (v)、能願動詞 (aux)、狀態動詞 (vs)、名詞 (n)、代名詞 (pron)、連接詞 (conj)、介詞 (p)、數詞 (num)、指示詞/定詞 (det)、量詞 (cl)、句末助詞 (sp)、時態助詞 (asp)、副詞 (adv)、助詞 (de)、疑問詞 (que)、複數詞綴 (plural)。

(2) 「語」的層次 (11類)，如主語 (sub)、賓語 (obj)、名詞短語 (np)、動詞短語 (vp)、介詞短語 (pp)、修飾語 (mod)、時間詞語 (time)、地方詞語 (loc)、及物性 (tran)、[數/定+量]結構 (dm)、離合結構 (vo)。

(3) 句型層次 (7類)，如複雜名詞子句 (rel)、特殊句型 (把

ba、被bei、讓rang、是shi、有you)、其他句式(pattern)。

(4) 混合類(2類), 如形式錯誤(form)、表達方式錯誤(sentence)。

標記大原則

1. 學習者的偏誤表現, 應該是具系統性、具解釋性的, 不過在標記過程中, 有時很難判定是學習者的偏誤或失誤, 因此有時我們所標記的涵蓋「失誤」及「偏誤」。如下述例句(1), 無法確定是輸入時的失誤將「靜」打成「間」或是偏誤, 將「夜深人靜」誤用為「夜深人間」。

(1) 我並不反對你夜深人間時, 玩線上遊戲。

如果是明顯的打字或編輯問題, 不做錯誤標記, 而是直接訂正原文。例如「他給我跟我介姊姊紹了他的朋友」顯然是考生文書處理時(註: TOCFL是電腦輸入作文)的錯誤, 我們訂正原文為「他給我跟我姊姊介紹了他的朋友」即可。

2. 做病句修正時, 有兩大原則。(a) 盡量不改動作者想表達的意思; (b) 該階段學習者能力所具備的詞語(可能對目標語人士不是那麼自然, 但合乎語法)。

3. 標記時, 先決定主類, 再選擇次類。如例句(2), 「真」多餘, 因此是R類, 「真」是副詞, 因此選擇adv。

S(選用偏誤)的次類標記選擇則和其他主類不同, 如例句(3)「去」錯用, 應該是「要」, 此時次類標記應該選擇aux, 也就是S的次類是以應使用的正確成分為標記選擇。

(2) 他是真[Radv]很好的人。→他是很好的人。

(3) 李台生建議他不去[Saux]一個人爬山。→李台生建議他不要

一個人爬山。

4. 原則上，標記以詞為最小原則（除了「們」以外），因此像例句（4），要將「不良性」當成一個詞來看，而非標記「性」為多餘成分（註：如果不確定「詞」的切分，可上平衡語料庫查詢 <http://140.109.19.114/>）。例句（5）也是視為一個詞，詞間的成分不對，尤其是四字格問題，標記為選用錯誤，次類為形式問題form。

（4）身體立刻有不良性[Svs]的反應。→身體立刻有不良的反應。

（5）我並不反對你夜深人靜[Sform]時，玩線上遊戲。→我並不反對你夜深人靜時，玩線上遊戲。

5. 關於次類標記的選用是有優先順序的，以下是幾個原則：

A. 偏誤牽涉到特殊句式，也就是「把、讓、有、被、是」，不論是多用、少用、錯用、錯序，應優先標示該次類標記ba、rang、you、bei、shi，而非p、pp、v、vp……。

（6）我現在沒有時間寫下來所有要告訴你的東西[Mba]。→我現在沒有時間把所有要告訴你的東西寫下來。

（7）錄取那個課程的話[Mbei]。→被那個課程錄取的話。

（8）這個[Sdet]話的印象非常深刻[Mrang]→這些話讓我印象非常深刻。

但是，下述情況地方詞語或時間詞語的錯序標記還是優先於特殊句式：

（9）他在櫃子裡[Wloc]把衣服放。→他把衣服放在櫃子裡。

B. 在錯序、遺漏、冗贅的情況下，名詞或名詞短語針對它們在句中功能優先標示為主語或賓語，如例句（10）；如果是名詞錯用，

則標示n，如例句（11）；主語或賓語中遺漏了部分成分，則標記為np，如例句（12）。

（10）我慶祝會[Wobj]不能參加。→我不能參加慶祝會。

（11）然後從公司到等公車的地[Sn]。→然後從公司到等公車的地方。

（12）我們到現在都是以寫信為連絡[Mnp]。→我們到現在都是以寫信為連絡的方式。

C. 表示地方或時間的介詞短語、名詞或名詞短語，優先標示為time或loc。如例句（13）、（14）。而非標示n、np或pp。

（13）進名校是在韓國[Wloc]非常重要的關鍵。→進名校在韓國是非常重要的關鍵。

（14）明天在我家下午2點30分[Wtime]集合。→明天下午2點30分在我家集合。

插入標記位置的原則

1. M類：標記在該句缺少的偏誤成分的空缺處，如例句（1）。如果屬於句式漏用，則放句末，如例句（2）。

（1）我真的[Madv]緊張要跟林美美見面→我真的很緊張要跟林美美見面。

（2）我現在沒有時間寫下來所有要告訴你的東西[Mba]。→我現在沒有時間把所有要告訴你的東西寫下來。

2. R類：標記在該句多餘的偏誤成分之後，如例句（3）。如果屬於句式多用，則放句末，如例句（4）、（5）。

（3）我知道妳又很[Radv]聰明又用功。→我知道妳又聰明又用

功。

(4) 而且我覺得她男朋友對她這樣做是太過分的[Rpattern]。→而且我覺得她男朋友對她這樣做太過分。(標記在整個句式之後)

(5) 以免更多人被其食品中毒[Rbei]。→以免更多人食物中毒。(標記在句末)

3. S類：標記在該句選用錯誤的偏誤成分之後。

(6) 我還不[Sadv]決定要在哪一家餐廳吃。→我還沒決定要在哪一家餐廳吃。

(7) 我不會記得[Sform]到火車站怎麼走。→我記不得到火車站怎麼走。

(8) 下個星期什麼空我都[Ssentence]。→下個星期我都有空。(通常標記在句末)

(9) 如何可能你那麼絕佳的同學[Ssentence]。→你是一個如此聰明的同學。

(10) 然後他爬起山來了[Sform]。→然後他開始爬山。

4. W類：標記在該句出現語序偏誤的成分之後，如例句(11)。優先標記屬於狀語功能的成分，如例句(12)，標記要放在「一直」後面，不是放在「心裡」後面。其次是名詞(n)、主語(sub)、賓語(obj)，如例句(13)，不標記在「敬上」後面，而是「李希平」後面。會標記在動詞組(vp)後面情況的應該很少，大概是連動結構才會發生。

(11) 今天早上我起床六點半[Wtime]。→今天早上我六點半起床。

(12) 一直[Wadv]心裡覺得很痛苦。→心裡一直覺得很痛苦。

(13) 敬上李希平[Wsub]→李希平敬上

5. 兩個都是狀語功能的成分錯序時，標記在第二個成分之後。

(14) 今天他先要[Waux]做功課再到學校去上課。→今天他要先做功課再到學校去上課。

(15) 吃飽以後再要[Waux]運動一下吧！→吃飽以後要再運動一下吧！

(16) 明天在我家下午2點30分[Wtime]集合。→明天下午2點30分在我家集合。

6. VO結構：W類的標記，放在整個結構後面，如例句(17)。

M類的標記，放在VO動詞之後，如例句(18)、(19)。

(17) 他就停車了[Wvo]幫我→他就停了車幫我。(「了」要插在VO中間)

(18) 他走路[Mvo]了三個小時→他走路走了三個小時。(VO動詞要重複)

(19) 你很久以前找工作[Mvo]很辛苦→你很久以前找工作找得很辛苦。

偏誤標記舉例 (以次類字母排序)

一、詞的層次

1. 副詞錯誤 (adv)

(1) 他們看了新電影就[Madv]高興。→他們看了新電影就很高興。

(2) 你是真的[Wadv]很好的人。→你真的是很好的人。

(3) 安卡拉不[Sadv]那麼熱。→安卡拉沒那麼熱。

- (4) 然後他一直走一直走就天快要[Radv]黑了。→然後他一直走一直走就天黑了。
- (5) 所以你快一點[Sadv]找到了工作。→所以你很快地找到了工作。
- (6) 我差不多[Sadv]忘了告訴你什麼方式可以認得我。→我差一點忘了告訴你什麼方式可以認得我。

2. 時態助詞錯誤 (asp)

- (7) 公共汽車沒有空的椅子，所以兩個人都得站[Masp]。→公共汽車沒有空的椅子，所以兩個人都得站著。
- (8) 我朋友也找工作找了很久才找到了[Rasp]。→我朋友也找工作找了很久才找到。
- (9) 他爬山爬的[Sasp]半天。→他爬山爬了半天。
- (10) 我放假的時候去歐洲了[Wasp]。→我放假的時候去了歐洲。
(這時使用W，而不是使用R+M的方式標記，請參考例句(26))

3. 能願動詞錯誤 (aux)：能願動詞包括「要、想、願意、肯、敢、應該、應當、應、該、得 (dei3)、能、能夠、可以、可、得 (de2)、可能、會、打算」。

- (11) 你們得好好地念書才[Maux]上大學。→你們得好好念書才能上大學。
- (12) 你願意[Raux]幫我好不好？→你幫我好不好？
- (13) 有一部電影我很要[Saux]看。→有一部電影我很想看。
- (14) 吃飽以後再要[Waux]運動一下吧！→吃飽以後要再運動一下吧！

4. 量詞錯誤 (cl)

- (15) 我真的高興你找到一[Mcl]新的工作了。→我真的高興你找到一個新的工作了。
- (16) 還有休息幾個[Rcl]天。→還有休息幾天。
- (17) 今天我上了一個[Scl]數學課。→今天我上了一堂數學課。

5. 連接詞錯誤 (conj)

- (18) 我想說你要想在那個時間做那件事是正常[Mconj]不正常。
→我想說你要想在那個時間做那件事是正常還是不正常。
- (19) 要不然[Rconj]如果成績不那麼好大概會傷她的心。→如果成績不那麼好大概會傷她的心。
- (20) 不論[Sconj]你自己覺得是小事情。→雖然你自己覺得是小事情。
- (21) 我們一般來說[Wconj]能從那個捷運站走路或者坐公車。→一般來說我們能從那個捷運站走路或者坐公車。

6. 「的、得、地」錯誤 (de)

- (22) 今天我決定我也要當一個很好[Mde]人。→今天我決定我也要當一個很好的人。
- (23) 今天他只有數學的[Rde]課。→今天我上了一堂數學課。
- (24) 時間過了[Sde]真快。→時間過得真快。
- (25) 突然我的[Rde]另外[Mdm]朋友打電話給我說。→突然我另外一個朋友打電話給我說。
- (26) 用我今年的[Rde]夏天放假[Mde]時間。→用我今年夏天放假的時間。(這時不用W，而用R+M方式標記)

7. 指示詞／定詞錯誤 (det):包括「這、那、當(天)、別(人)、

另外、該（地）、這些、那些、有些、一點、一些、整（個）、每（人）、其他、其它、隔（一天）、第、某、上、下、全、全部」。

- (27) [Mdet]一天很有意思。→這一天很有意思。
- (28) 我會在第[Rdet]一樓等你。→我會在一樓等你。
- (29) 我可以給你那麼[Sdet]錢。→我可以給你那些錢。
- (30) 也要花有些[Sdet]錢。→也要花一些錢。
- (31) 我們認為一邊可以多知道[Mdet]中華文化，……。→我們認為一邊可以多知道一點中華文化，……。

8. 名詞錯誤 (n)

- (32) 因為泰國菜是世界上最美味佳餚[Rn]的食物。→因為泰國菜是世界上最美味的食物。
- (33) 回家的時候，天氣[Sn]已經黑了。→回家的時候，天已經黑了。
- (34) 這裡什麼式[Sn]的食物都有。→這裡什麼種類食物都有。

9. 數詞錯誤 (num)

- (35) 是愛吃有機蔬菜的[Mnum]個消費者。→是愛吃有機蔬菜的一個消費者。
- (36) 小王是一個兩[Snum]年級的高中學生。→小王是一個二年級的高中學生。
- (37) 你不必全一[Rnum]天戴著。→你不必全天戴著。

10. 介詞錯誤 (p)

- (38) 你這樣子做不但[Mp]你的健康好。→你這樣子做不但對你

的健康好。

(39) 我們跟[Rp]一起坐捷運。→我們一起坐捷運。

(40) 我想給[Sp]你直接說「恭喜！」→我想跟你直接說「恭喜！」

(41) 感動了[Sp]沒話說。→感動到沒話說。

11. 代名詞錯誤 (pron)：為了和sub, obj區分，如果是M、W類的優先給sub, obj；如果是S、R類（通常是指稱錯誤或回指問題）則給pron。

(42) 所以要見面之前就知道互相[Spron]的個性。→所以要見面之前就知道彼此的個性。

(43) 我們應該[Mdet]一個人練習說他們[Rpron]自己的部分。→我們應該每一個人練習說自己的部分。

12. 疑問詞錯誤 (que)：指「什麼、怎麼、幾、怎麼樣、哪些、為什麼、哪裡」。

(44) 開始慢慢地教她這麼[Sque]彈鋼琴。→開始慢慢地教她怎麼彈鋼琴。

(45) 可是什麼[Wque]買？→可是買什麼？

(46) 在生活上你碰到[Mque]困難嗎？→在生活上你碰到什麼困難嗎？

13. 句尾助詞錯誤 (sp)

(47) 你可以告訴我最近的火車站在哪裡[Msp]？→你可以告訴我最近的火車站在哪裡嗎？

(48) 你可不可以送我到火車站嗎[Rsp]？→你可不可以送我到火車站？

(49) 如果你有空，我想再開一個好吧[Ssp]？→如果你有空，我想再開一個好嗎？

(50) 您可不可以跟我們一起來參加好嗎[Rsp]？²→您可不可以跟我們一起來參加？

14. 動詞錯誤 (v) (包含動詞補語錯誤，視動補結構為一個詞。這裡採用中研院斷詞規則。要注意動詞+時態助詞「了、著、過」是切分的，與中國大陸標準不一樣)。

(51) 謝謝您請我[Mv]你的慶祝會。→謝謝您請我參加你的慶祝會。

(52) 他們打算一塊兒在101大樓去[Rv]看電影。→他們打算一塊兒在101大樓看電影。

(53) 下[Sv]了電影我就回家了。→看完了電影我就回家了。

(54) 他一看[Sv]小安的汽車就放心了。→他一看到小安的汽車就放心了。

15. 狀態動詞錯誤 (vs)

(55) 他的方法不太溫柔，可是我在很短的時間內進步得很[Mvs]。→他的方法不太溫柔，可是我在很短的時間內進步得很快。

(56) 以前已經給你我的大[Rvs]電話號碼。→以前已經給你我的電話號碼。

(57) 日本的天氣非常清楚[Svs]。→日本的天氣非常清爽。

² 這裡多餘的是「好嗎」，應該是「語」的層次了，但考量到多數Sp都是詞層次，目前歸類在「詞」中。

16. 「們」錯誤 (plural)

(58) 不過我希望我們會變成好朋友們[Rplural]。→不過我希望我們會變成好朋友。

(59) 所有的同學們[Rplural]都是我的戰友。→所有的同學都是我的戰友。

二、語的層次

1. 地方詞、方位詞錯誤 (loc)：這個標記包括詞與語的概念。如例句(60)～(63)是「上、裡」等詞語問題；(64)則是「在日本」這個詞組的問題。

(60) 考試上[Rloc]的時候。→考試的時候。

(61) 李台生就去火車站裡[Rloc]坐公車回家。→李台生就去火車站坐公車回家。

(62) 要是你在功課裡[Sloc]有問題→要是你在功課上有問題。

(63) 我要帶你去在同安街[Mloc]我常常去的餐廳。→我要帶你去在同安街上我常常去的的餐廳。

(64) 如果我買那件衣服在日本[Wloc]。→如果我在日本買那件衣服。

2. 修飾語錯誤 (mod)：名詞前的修飾語屬於〔X+的〕結構，標記為mod。

(65) 現在我知道甚麼樣的[Wmod]買東西是值得的。→現在我知道買甚麼樣的東西是值得的。

(66) 因為你要到我的城市來看看[Mmod]生活怎麼樣。→因為你要到我的城市來看看這裡的生活怎麼樣。

(67) 我希望你改掉這兩個你的[Rmod]生活習慣。→我希望你改

掉這兩個生活習慣。

3. 名詞短語錯誤 (np) (包括「的+N」的偏誤)

(68) 我跟同學一起住在房租[Sn_p]。→我跟同學一起住在租的房子。

4. 賓語錯誤 (obj)

(69) 他們說那個電影[Wobj]想看。→他們說想看那個電影。

(70) 他請我教[Mobj]日文。→他請我教他日文。

(71) 我不知道你可不可以幫我拿回來一瓶紅酒[Wobj]。→我不知道你可不可以幫我拿一瓶紅酒回來。

5. 介詞短語錯誤 (pp)

(72) 他坐在教室前跟林美美[Wpp]。→他跟林美美坐在教室前。

(73) 希望你在太晚的時候[Rpp]早點去睡。→希望你早點去睡。

(74) 不可以買給別人[Wpp]飯→不可以買飯給別人。

6. 主語錯誤 (sub)

(75) 我聽說[Msub]打算到法國。→我聽說你打算到法國。

(76) 害得[Msub]全家食物中毒。→害得我們全家食物中毒。

(77) 所以決定我[Wsub]一個人去。→所以我決定一個人去。

7. 時間詞語錯誤 (time) : 這個標記包括詞與語的概念。

(78) 可是現在[Rtime]最近我工作很忙。→可是最近我工作很忙。

(79) 他們正在[Stime]還在電影院一邊看電影一邊聊天。→他們現在還在電影院一邊看電影一邊聊天。

(80) 他平常起床七點鐘[Wtime]。→他平常七點鐘起床。

8. 及物性錯誤 (tran) : 不及物動詞後面加了賓語，標記為tran。

(81) 有三個日本學生入學[Wtran]了我們的學校。→有三個日本學生在我們學校入學。

(82) 見面[Wtran]她的老公時，……。→跟她的老公見面時，……。

(83) 我實在不好意思[Wtran]父母。→我實在對父母不好意思。

9. 動詞短語錯誤 (vp)

(84) 他們去學校前的公車站等一等[Svp]。→他們去學校前的公車站等公車。

(85) 有問題的話，他們可以到我們家來打電話[Wvp]。→有問題的話，他們可以打電話到我們家來。

10. 數量詞錯誤 (dm)

(86) 他還買給自己[Mdm]筷子。→他還買給自己一雙筷子。

(87) 因為我們班有各種[Rdm]各國的同學們[Rplural]。→因為我們班有各國的同學。

(88) 然後去一個[Wdm]台北附近的山。→然後去台北附近的一座山。

11. 動賓結構離合所造成的問題 (vo) : 可能是中插成分的位置問題，如例句 (89)、(90)；或是動詞重複的問題，如例句 (91)、(92)。

(89) 可是一次也沒見面過[Wvo]。→可是一次也沒見過面。

(90) 寫信也很多[Wvo]。→也寫了很多信。

(91) 因為我們在網路聊天[Mvo]了快一年了。→因為我們在網路聊天聊了快一年了。

(92) 她賺錢[Mvo]得很辛苦。→她賺錢賺得很辛苦。

三、句型層次

1. 「把」字句錯誤 (ba)：這個標記包括詞與句型的層次。有時指的是「把」這個詞，有時指的是把字句。如果是該用把字句而沒用 (Mba)，或不該用把字句而用了 (Rba)，也就是不是單單一個「把」的問題，標記要放在句末。如例句 (96)。

(93) 先去飯店[Mba]行李放在床上。→先去飯店把行李放在床上。

(94) 我希望你每天都讓[Sba]我們的房間弄乾淨。→我希望你每天都把我們的房間弄乾淨。

(95) 之後把[Rba]我的中文能力也很快地進步了。→之後我的中文能力也很快地進步了。

(96) 下課以後常常把學校的事一起討論[Rba]。→下課以後常常(跟他)一起討論學校的事。

2. 「被」字句錯誤 (bei)：這個標記包括詞與句型的層次。有時指的是「被」，有時指的是被字句。標記原則如上述 (把字句)。

(97) 她的臉被[Rbei]受傷流血。→她的臉受傷流血。

(98) 每一個人都需要有一些好朋友被[Rbei]幫他們過人生的路。
→每一個人都需要有一些好朋友幫他們過人生的路。

3. 「讓」字句錯誤 (rang)：這個標記包括詞與句型的層次。

(99) 不但給自己開心也[Mrang]大家看得很快樂的。→不但給自己開心也讓大家看得很快樂。

(100) 但是我的老師馬上就把[Srang]我心裡安靜下來。→但是我的老師馬上就讓我心裡安靜下來。

4. 「是」錯誤 (shi)：這個標記包括詞與句型的層次。

(101) 下個星期五[Mshi]我弟弟的生日。→下個星期五是我弟弟的生日。

(102) 小明看不出來老師是[Rshi]50歲。→小明看不出來老師50歲。

(103) 我覺得一件白色跟藍色格子的褲子想[Sshi]很好的禮物。→我覺得一件白色跟藍色格子的褲子是很好的禮物。

(104) 我們看的電影好像[Mshi]最有名的。→我們看的電影好像是最有名的。(這裡標記shi，而非pattern，可參考例句(111))

5. 「有」錯誤 (you)

(105) 平常也不太常有[Ryou]下雨。→平常也不太常下雨。

(106) 因為北海道是[Syou]很多好吃的東西和漂亮的風景。→因為北海道有很多好吃的東西和漂亮的風景。

(107) 她還說學習的路很長[Myou]很多困難。→她還說學習的路很長有很多困難。

6. 複雜名詞子句錯誤 (rel)：這一類只出現在S主類。大部分是關係子句的表達問題，以及名詞中心語 (Head) 位置錯誤。

(108) 我覺得餐廳在你家旁邊[Srel]好像好吃。→我覺得在你家旁邊的餐廳好像很好吃。

(109) 希望我們的計畫要見面[Srel]沒有甚麼問題。→希望我們要見面的計畫沒有甚麼問題。

(110) 我覺得野餐是最好的辦法認識新朋友[Srel]。→我覺得野餐是認識新朋友的最好辦法。

7. 其他句式錯誤 (pattern)：包括「是……的、連……也、除了……以外、一邊……一邊……、」等有框架形式的句型。標記插入在框架之後，如果沒能找到框架，大部分放在句末。

(111) 我這麼覺得[Mpattern]。→我是這麼覺得的。

(112) 我在書面上[Spattern]已經了解你的個性。→我從書面中已經了解你的個性。

(113) 他告訴那個人是他電話號碼的[Rpattern]。→他告訴那個人他的電話號碼。

四、混合類

1. 形式錯誤 (form)：主要有三種情況

a. 選用的詞沒錯，只是形式錯誤（例如重疊式、嘗試貌），如例句

(114) ~ (116)。

b. 可以理解語義，只是表達形式錯誤，如例句 (117) ~ (119)。

（可能造出的不是中文詞）

c. 遺漏了固定格式中的某部分，如例句 (120)。

(114) 上課的時候我跟愛文常常談一談[Sform]不知道的問題。→上課的時候我跟愛文常常談不知道的問題。

(115) 他在路上走著想想[Sform]他的功課。→他在路上走著想他的功課。

(116) 誰都不知明天也[Radv]自己的工作或公司還會[Sform]存在。→誰都不知明天自己的工作或公司還會不會存在。

(117) 幸運[Mform]我住的地方有越南餐廳。→幸運的是我住的地

方有越南餐廳。

(118) 李文平敬於[Sform]。→李文平敬上。

(119) 我覺得不舒服得不得了[Sform]。→我覺得非常不舒服。

(120) 可是對他們[Mform]是開心的一天吧！→可是對他們來說是
開心的一天吧！

2. 句子表達錯誤 (sentence)：當學習者的句子中不只出現一個錯誤，母語者雖然能理解其意思，但該句錯誤太多，需要做整體性的修改時，標記為[Ssentence]。或是連續三個（含）以上的錯誤，即使可以歸類為詞層次偏誤，仍以sentence來標記。

(121) 合格了交換學生[Ssentence]以後。→符合了交換學生的資格以後。

(122) 她是最好的家教一樣的感覺[Ssentence]。→感覺她是最好的家教一樣。