

壹、緒論

在這個資訊爆炸的時代，巨量資料（big data）已成為一股新興的研究趨勢，如何從巨量資料中擷取其富含的珍貴訊息也成為研究者高度關注的議題。巨量資料的研究方法近年來也大量被應用在外語學習的研究中，各國學者非常積極地蒐集大量外語學習者的資料並建置資料庫，以用來探討外語學習的歷程（Granger, 2009）。巨量資料的研究典範將語言學習帶進了一個新的紀元，使外語學習由過去以小樣本為主要的研究方式邁進更大量、更具代表性的資料探討學習歷程的研究趨勢（Granger, 2009; Myles, 2005）。當前的資料庫多數為提供外語學習者用的學習材料，內容多數為口語或書面短文、新聞、聊天等學習用的題材，或是記錄外語學習者所產出的文章、句子，以用來檢驗外語學習者的錯誤形態，針對大量詞彙蒐集外語學習者反應時間來探討單一字詞辨識歷程的資料庫仍相當缺乏。此外，將資料庫中的大量資料進行系統性分析的研究也仍鮮有（Granger, 2009）。有鑑於此，本研究將延續先前的研究成果，試圖建立並擴充臺灣地區英語為外語學習者大學生做為對象的心理詞彙資料庫，現階段將使用此資料庫來檢驗臺灣地區大學生的英語字詞辨識歷程，未來更期望能提供其他研究者探討英語學習

者閱讀研究與教學的刺激材料之參考依據。

目前，國內、外有相當多針對外語學習者建置的資料庫，例如由德國 University of Bremen 為了幫助學生撰寫學術性文章，蒐集大量英語學習者所撰寫的短文、摘要、計畫書等學術性文章，並建置成學術英語語料庫（Corpus of Academic Learner English, CALE）（Callies & Zaytseva, 2013），該語料庫不僅提供英語學習者常犯的錯誤類型，更透過分析巨量資料獲得的結果提供學習者自身的英語程度指標；類似性質的語料庫還有上海交通大學與廣東外語外貿大學針對中國的英語學習者共同建置的英語學習者語料庫（The Chinese Learner English Corpus, CLEC）（Gui & Yang, 2003）、瑞士 University of Uppsala 針對一般學生設計的學生英語語料庫（The Uppsala Student English Corpus, USE）（Axelsson, 1999; Axelsson & Hahn, 2001）、比利時 Université catholique de Louvain 針對不同母語之中高階英語學習者設置的國際英語學習語料庫（The International Corpus of Learner English, ICLE）（Granger, 2003）等。在國內也有成功大學外文系彙整日、西、德、中文之語言學習材料所設置之多國語言學習語料庫（<http://corpora.flld.ncku.edu.tw/>），師大也為華語學習者設置華語學習語料庫

(<http://kitty.2y.idv.tw/~hjchen/cwrite/main.cgi>)，目前已收錄300萬字。這些語料庫為外語學習者提供相當完善的學習材料，使用巨量資料的訊息提供學習者回饋並幫助語言學習已儼然成為一股銳不可當的新趨勢。然而，當前的資料庫多數為針對受試者的錯誤樣態進行分析，目前仍鮮少有研究以巨量資料方法收錄外語學習者進行單一字詞反應時間之語料庫，因此這類型的資料庫尚待建置。

在閱讀的研究中，反應時間與正確率是用來探討字詞辨識歷程相當重要的指標 (Katz et al., 2012)，因此目前有許多國家針對不同語言之母語使用者蒐集了反應時間的巨量資料，例如由美國六所大學共同建置的英語詞彙計畫 (English Lexicon Project)，其中收錄包含815名受試者針對40,481個真詞與假詞的詞性判斷 (lexical decision task) 與唸名 (naming) 反應時間 (Balota et al., 2007)；比利時Ghent University也有進行荷語、英語、法語詞彙計畫，其中收錄各種語言母語使用者之反應時間，以及習得年齡、詞彙具體程度、情緒等主觀判定 (Brysbaert, Warriner, & Kuperman, 2014; Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010; Keuleers, Lacey, Rastle, & Brysbaert, 2012; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012;

Moors et al., 2013; New, Pallier, Brysbaert, & Ferrand, 2004; Warriner, Kuperman, & Brysbaert, 2013)。在外語學習的研究方面，目前僅有張葦葦與李俊仁 (2009) 以紙本測驗方式蒐集120名臺灣地區大學生對於3,075個英文詞的辨識度、熟悉度評定，與習得年齡，但受限於使用量表的研究方式，該研究並無受試者對於這些英文詞的反應時間資料。有鑑於此，本研究使用與張葦葦與李俊仁相同的刺激材料，但改以電腦施測，邀請受試者對3,000個英文詞進行詞彙判斷作業，並記錄其正確率與反應時間，除擴充該資料庫的內容，也希望更完整地探討臺灣地區英語學習者字詞辨識歷程的機制。

在探討英語為母語者的閱讀歷程研究中，詞長效應 (word length effect) 是一個經常被用來檢視字詞辨識歷程的重要指標 (Cohen et al., 2003; Ferrand & New, 2003; Frederiksen & Kroll, 1976; New, Ferrand, Pallier, & Brysbaert, 2006; Weekes, 1997; Whaley, 1978)。詞長效應指的是隨著一個詞的長度愈長，受試者辨識時間也隨之增加的現象，進行字詞辨識時若出現詞長效應，一般被認為顯示受試者可能採取序列式的方式，將詞依照形或是音的最小單位做分割之後進行逐片段處理，這種處理方式稱為次詞彙 (sublexical) 層次處理，這是由於此時字詞辨識的處理