

## 影響中文文本可讀性之句子難度因素探究\*

洪嘉駝\*\*

國立臺灣師範大學華語文教學系

陳家妮

國立臺灣師範大學華語文教學系

### 摘要

近年來，中文閱讀理解能力的重要性逐漸受到重視。主要是因為中文閱讀理解能力的影響層面很廣，不論是在以中文為母語的習得或以中文為二語的習得的過程中，閱讀理解都被視為重要的能力展現之一。關於閱讀理解能力的影響，從詞彙或語法至篇章，皆涉及句子的難度與複雜程度（Hong et al. 2020；唐玉玲、于東 2020），因此在中文作為二語的教學中，組成句子的詞彙與語法常作為判斷閱讀能力程度的標準（Silberstein 1987），然而，對於中文句子難度的相關研究，多僅以單一指標來探討，有的研究僅討論詞彙特徵的複雜性（Ye and Qiu 2008；蔡美智 2010；陳旻祺 2013；陳司恩 2019）、有的研究僅討論特定單一語法句式的多樣性（張莉萍 2005；肖奚強等人 2009；陳珪 2010；王昱勻 2021），缺乏具整合性的客觀量化標準，以致於對於量化句子難度、選擇特徵量化句子難度，尚未得到共識（鄭錦全 2005；Ye and Qiu 2008；龐成 2016；吳思遠等人 2020）。

因此，本研究以語料庫語言學為本，從中文文本的角度，透過中文語法數位平台探究詞彙複雜性與語法複雜性對中文句子難度的影響，並以量化訊息層面進行詞彙與語法的難度排序，再以華語八千詞表、文本可讀性指標自動化分析系統以及國際中文教育中文水平等級標準進行驗證，比對各項語法之難度，進而建立具系統性及客觀性的中文語法點難度排序，以輔助中文作為二語的句子教學，提升中文為二語之學習者的句子理解能力與閱讀能力。

**關鍵詞：**文本可讀性 句子難度 詞彙複雜度 語法複雜度 閱讀理解

---

\* 本研究獲國科會（MOST 110-2511-H-003 -034 -MY3）經費補助及教育部核定之國立臺灣師範大學「高等教育深耕計畫」經費補助。感謝本期刊兩位匿名審查者提供寶貴建議，使本文更臻完善。文中如有疏漏，悉由作者負責。

\*\* 本文通訊作者。

## 1. 前言

句子是能夠清楚表達的基本單位（龐成 2016），亦是掌握閱讀理解的關鍵，而組成句子的成分包含詞彙與語法，在學習評量的評判上，常以句子中的詞彙與語法作為評判學習者閱讀能力程度的標準（Silberstein 1987）。

於教學而言，無論是對以中文為母語的學習者或是以中文為二語的學習者，閱讀理解皆被納為重要的語言能力指標（宋曜廷等人 2013；吳思遠等人 2020），且許多研究亦證實閱讀理解與句子難度息息相關（Hong et al. 2016；Hong et al. 2020；唐玉玲、于東 2020），然而有關中文句子難度的研究當中，有部分研究以詞彙特徵作為句子難度指標（Tsao 1996；Ye and Qiu 2008；楊惠媚等人 2015），亦有研究以特定語法類型進行討論（張莉萍 2005；尚奚強等人 2009；Hong et al. 2016），無論是採用詞彙特徵還是以特定語法劃分，語料來源不一，使得研究結果因而有所不同，在二語教學的使用上，也因此產生多種難度指標，難以客觀鑑定所有教材的難度以及中文作為二語之學習者的程度；此外，過去研究多偏向質化討論，對於如何量化句子難度，以及究竟要選擇什麼特徵以量化句子難度尚未得到共識（鄭錦全 2005；Ye and Qiu 2008；龐成 2016；吳思遠等人 2020），可見關於詞彙和語法對中文句子難度影響的相關研究仍有發展空間，建置完善且客觀的句子難度指標亦為迫切需求。

基於前人研究中針對各項語法點在語義、句法、語用等方面的深入了解，並證實了詞彙複雜性與語法複雜性確實會影響句子難度與理解，因此本研究將在前人的研究基礎上，了解詞彙複雜性與語法複雜性對句子難度的具體影響，將詞彙與語法加以分析並整合，建立具系統性及客觀性的中文句子難度指標，以期能輔助中文作為二語的句子教學，有效協助學習者提升閱讀能力。

## 2. 研究目的

許多研究皆指出詞彙和語法特徵都將影響句子難度，但多數研究僅針對詞彙複雜性或單一語法進行探討（Laufer and Nation 1995；Ye and Qiu 2008；Hong et al. 2016；王進等人 2017），即便有研究同時納入詞彙和語法，但受到研究實驗的設計影響，仍無法判斷詞彙和語法對於句子難度的影響為何（Diana et al. 2011）。

在針對華語教學的相關研究中亦有相同的狀況，詞彙複雜性的相關研究

常為了明確列出學習者應學之詞彙，會依據詞彙等級、情境等製作詞表或建置語料庫。楊惠媚等人（2015）認為常使用的詞表如《華語八千詞表》，是參考書面語語料庫而製成，然而學習者的口語交際需求一般比書面交際的需求還要高，因此於該研究中以口語語料庫中的高頻詞對比華語八千詞表的詞彙，並提出適合加入詞表的詞彙；而針對語料庫，孫雪芹（2023）也指出各類型語料庫雖然都能因應課程需求作為教學的材料或參考依據，但語料庫之間的來源不同，容易出現超出學習者等級的詞彙，而各語料庫對於詞彙的分級，由於採用之標準不同，也可能出現落差。另一方面，有關華語單一語法的研究，如張莉萍（2017）以學習者語料庫為本，從偏誤句分析母語為英語之學習者與母語為日語之學習者學習關係子句的情形；吳默（2021）則從教學的角度出發，探討「並」作為動詞、副詞以及連詞時語用功能的異同，以及產生教學難點的原因，並提出可行的教學策略。

由於各研究所選取的指標不一致，導致研究成果大相逕庭，難以統整出具規範性且客觀的語法點難度標準，實際應用至中文作為二語教學時，亦會因此出現判斷結果不一致的現象。故在已知句子難度會受到詞彙和語法複雜性影響的情況下，在建立句子難度的時候應考量這兩種複雜性特徵，不應以經驗導向偏好某一項複雜性特徵的使用，綜上述而論，本研究所要探討的方向為探討詞彙複雜性與語法複雜性對中文句子難度的影響，並建立客觀的中文語法點難度排序。

因此，本研究透過蒐集中文句子難度的相關文獻（McLaughlin 1969；鄭錦全 2005；Ye and Qiu 2008），以「中文語法數位平台」及「華語八千詞表」為依據，進行分析並加以整合，確認影響句子難度的詞彙指標與語法指標，以建立具系統性與客觀性的句子難度指標，以期能輔助句子教學與文本教學，提升學習者的閱讀能力。

### 3. 文獻探討

句子是組成篇章的基本單位，句子難度即為評估句子能被理解的程度，所以在探討閱讀理解的相關文獻中，不乏有從句子做為切入討論角度的相關研究（Klare 2000；Diana et al. 2011；Lu and Ai 2015；龐成 2016；Lu et al. 2019；吳思遠等人 2020）。

有關句子難度的量化研究，過去許多學者考慮句子長度和句子成分的長度（Gunning 1952；McLaughlin 1969；Brown 1971；Klare 2000；Lu 2010；吳

思遠等人 2020)，也有許多研究是綜合考量詞彙和句子結構的複雜性( Spache 1953; Thoma and Gilbert 1967; Kenneth and Susan 1973; Laufer and Nation 1995; 鄭錦全 2005; Ye and Qiu 2008; Diana et al. 2011)，但關於詞彙和語法複雜性對於句子難度的影響程度尚未有明確相關研究，即便認同兩者都會影響句子難度，仍欠缺相關量化研究以致無法有效確認兩者對句子難度的影響，進而區分語法點之間的難度差異，故本節將由回顧句子難度的發展應用，進而深入探討詞彙複雜性與語法複雜性對中文句子難度的影響。

### 3.1 句子難度發展

關於句子難度的整體發展，從句子難度指標的建立到其應用，相關研究在中文和英文都已有一定的發展，且從過去的研究中發現，詞彙複雜性和語法複雜性對句子難度皆具有一定影響力( Spache 1953; Thoma and Gilbert 1967; Kenneth and Susan 1973; Laufer and Nation 1995; 鄭錦全 2005; Ye and Qiu 2008)。

早期有關於句子難度方面的研究常聚焦於不同年齡層的比較。Gaer (1969) 透過直述句、被動句、否定句，以及疑問句四類英語句型的使用和複雜性觀察 3 到 6 歲的兒童與成人理解句子和自行產出句子的能力。實驗結果發現，句法複雜性和心理複雜性是有關係的。因此，成人受試者使用具有嵌入子句的句子比例很高，但是複雜的句子，也會因為人的經驗和心理變化而有不同的結果。

Eisenberg (2002) 則研究 16 位 3 至 4 歲且英語為母語的兒童在聽力與閱讀能力測驗上，對句子理解的狀況，結果發現英語連接詞越多的句子表現普遍越差，尤其在閱讀測驗中，兒童經常仰賴前後文的邏輯去推演事件發展，但要是句子跳脫出線性或是時間性的邏輯，就會出現理解困難，因此比起詞語的影響，句法對於句子理解的影響是更顯著的。Kidd (2003) 針對 Eisenberg (2002) 的研究結果進行剖析，認為兒童從 2 歲開始就會產生語言結構，不過尚不穩固，所以不應單憑提供句子來研究兒童的理解能力，唯有給予更精準實務的條件，如針對英語關係子句的結構，才能對兒童的理解能力做出更精準的評估。

Gordon 與 Johnson (2004) 則研究在英語句子語義沒有歧異的狀態下，探討主從句的複雜程度觀察大學生對不同難度主從句的理解。其研究結果顯示，句子中的主語（是個人或是概括的代詞，如他們、大家等等）與敘述類

型，以及主語和賓語的關係會影響都會影響受試者的解讀判斷，而且不同類型的名詞子句干擾受試者對事件的記憶、思考與表達，也就是說，基於人們會有習慣使用的句式，腦中的記憶和思考習慣會使句法複雜句子的處理難度增加。

由此可見，詞彙複雜性與語法複雜性是句子難度的評判的關鍵，但缺乏規範性，為了更具科學性、邏輯性的檢視句子難度，相關研究開始討論語法點難度指標的建立 (Diana et al. 2011; Covington et al. 2006)，以及指標檢驗 (Lu 2008; Lu and Ai 2015)。

Rosenberg 與 Abbeduto (1987) 根據兒童語言發展的研究建立了 7 個等級的英語句法習得順序，其習得順序指標包含了詞彙複雜性、句子結構以及句子複雜性，研究結果發現受試者輸出複雜句子的能力確實與習得順序有關。Covington 等人 (2006) 對 Rosenberg 與 Abbeduto (1987) 提出的分級進行優化，從詞彙複雜性、詞彙型態變化、句子結構、句子複雜性以及心理語言學的實際測試等再細分為 8 個等級。接著，Lu (2008) 應用 Covington 等人 (2006) 優化的 D-level 分級嘗試建立有效的自動複句分析系統，並擴展了分析數量及範圍。基於先前對於 D-level 分級的分析，Lu (2010) 參考二語習得發展研究，得出 14 種評量方式設計一套自動分析系統，以二語寫作文本做為輸入來源 (input) 驗證其系統於陌生的二語寫作文本，並得到高信度 (high reliability)。

隨著句子難度相關研究的關注度提升，詞彙複雜性以及語法複雜性在第二語言 (L2) 教學中的重要性也日益顯著，為建立有效之句子難度分級，眾多研究聚焦於詞彙複雜性或語法複雜性進行探究，因此本研究接下來將分別針對詞彙複雜性與語法複雜性進行討論。

### 3.2 詞彙複雜性

詞彙為母語者及學習者接觸到的最小具意義之獨立單位 (竺家寧 1999)，因此詞彙複雜性的研究層面，從定義何為詞彙，比較詞彙的使用頻率，到統整詞彙的義項數量等等，皆為有關語言學習及教學研究所關注的重點。

Kenneth 與 Susan (1973) 以英語為基礎，比較詞彙 (word)、非詞彙 (nonword) 以及不常用詞彙 (unfamiliar word)，透過實驗發現受試者判斷一詞是否為詞彙時，所需時長和搜尋過程有關。該實驗結果指出，唸名時長 (naming time) 是詞彙搜尋過程的結果，為了辨認目標物是否為詞彙，會回

到長期記憶中尋找已儲存的資訊，找到後才完成為詞彙命名的過程，所以，越快搜尋到詞義所在的位置，過程就越快完成。因此，詞彙所需的時間比非詞彙短，且高頻詞的時間比低頻詞短。

Laufer 與 Nation (1995) 認為詞彙複雜性 (Lexical Complexity) 涉及到詞彙知識的廣度和深度，其測量指標主要可分為四類：(1) 詞彙複雜性 (lexical sophistication)，即文本中低頻詞占總詞數的比例；(2) 詞彙多樣性 (lexical diversity)，即詞類 (type) 與單一詞彙出現的次數 (token) 之間的比例；(3) 詞彙密度 (lexical density)，即文本中實詞占總詞數的比例；(4) 詞彙獨特性 (lexical originality)，即文本中的獨有詞 (於某個文本中使用，但其他比較文本皆未使用的詞) 占總詞數的比例，因此認定詞彙的難易程度是影響閱讀理解的基本因素之一。

Diana 等人 (2011) 的研究針對英語為二語的三年級學生在科學課文閱讀設計四個主題，結果發現，英語的詞彙複雜性對學生對四個主題中其中兩個主題的理解產生重大影響，推測如此的結果可能來自其中兩個主題的生詞在日常生活中常接觸，例如在「牙膏」的主題中，細菌 (bacterium) 在教科書中應是低頻詞彙，但因學生在生活中前往診所或醫院的時候較容易接觸該詞，提高學生認識這個詞的機會，便構成學生的先備知識之一。

Lu 等人 (2019) 則在分析科學寫作與科學影響之間的關係時，綜合詞彙以及句法的難度作為評比，其中詞彙難度考慮了詞彙的多樣性、精密度和詞彙使用密度，且該研究中發現科學寫作包含更多含有修飾語的名詞短語，這使句子在結構上更加緊密和抽象。

關於中文的詞彙複雜性研究中，鄭錦全 (2005) 針對中文句子難度提出三個影響因素，包含句子長度、詞頻和詞彙語義，認為詞彙的語義類別數目會影響句子的難易程度，當句子中語詞數目與語義類別比率越高，讀者越難讀懂，所謂語義類別指單詞在語義上的分類，多義的詞語較容易導致句子在語義層面上的歧異；另外，通常詞彙的語義類別數量愈多，詞彙語義變化越大。

為符合當前的華語教學需求，在接下來的發展中，許多研究僅聚焦於單一類型的詞彙研究。蔡美智 (2010) 依據漢語水平考試 (HSK) 的標準及相關書籍，探究詞性為名詞、動詞、形容詞及副詞的近義詞，並透過中高級學習者得偏誤情形，加以分析教學策略，認為在使用頻率的基礎上，當有近義詞出現時，應先提供句法功能，再以搭配詞輔助學習；陳旻祺 (2013) 分析

了語氣副詞「幸虧」、「幸好」、「好在」與「多虧」的語義、句法以及語用；陳司恩（2019）透過中研院現代漢語平衡語料庫之語料，針對「改正」、「糾正」、「矯正」等三個近義詞進行分析。

從過去的研究中，可見影響句子難度的種種因素中，詞彙佔有一定份量的影響度，陌生的詞彙或是難度過高的詞彙都會影響學習者的閱讀理解（Kenneth and Susan 1973；Laufer and Nation 1995；鄭錦全 2005；Diana et al. 2011；Lu et al. 2019）。所謂陌生詞彙即會考慮到詞頻、詞彙在單篇文本中出現次數以及學習者的先備知識，而詞彙語義數和詞組結構也會影響到閱讀理解（鄭錦全 2005；Lu et al. 2019），但這些研究的指標選擇不一，尚缺乏系統性。

為有條理地整合詞彙，Ai（2022）建置中文詞彙複雜程度在線分析工具（WeCLECA），辨析語料庫（Lancaster Corpus of Mandarin Chinese）中學術論文、小說與新聞等三種不同語體中的詞彙複雜性，認為通過分析文本的詞彙複雜性，可以協助中文為二語的教學找到更合適的教學材料，或作為檢視學習者學習狀況的工具，但同時，Ai（2022）也表示，該分析工具缺乏語法分析，對於了解學習者的語言學習掌握度是有限的，因此接下來將針對語法複雜性對句子難度的影響進行分析。

### 3.3 語法複雜性

語法是清楚表達想法的關鍵，亦常作為學習者語言能力及程度的評判標準，因此相關研究多以偏誤與閱讀理解兩個方面切入。Thoma 與 Gilbert（1967）探討錯誤的句法結構對句子閱讀理解的影響，主要是針對詞語位置錯置和錯誤的句法形式作討論，錯置的句子包含主語、主要動詞，以及賓語的錯置。該研究實驗以偏誤句混合正確句，參與人員看到句子後要立即說明句子語義，句子會持續出現直到說出正確的句子順序和語義，紀錄受試者的判讀時間。該研究結果顯示，錯誤的句子普遍會使閱讀時間增長，但在形容詞或副詞的錯序句子中最為明顯，若是遇到句法越複雜的句子，所需要的時間也相對增長。

Bowey（1986）在研究中探討文字解碼能力和閱讀理解能力的關聯，透過兩個假設：（1）不熟練的讀者（以文字解碼能力區分）其句法意識（syntactic awareness）比起較熟練的讀者來得差（2）句法意識與持續的閱讀理解監控和整體閱讀理解有關。其研究結果認為，閱讀理解監控需綜合評估所獲訊息的

句法和語義的一致性，讀者的文字解碼能力確實會影響其閱讀理解監控的結果，但並非唯一的原因；句法意識與閱讀能力確實有關聯，但不足以證明熟練的讀者比起不熟練的讀者有較良好的句法意識，句法意識與文字解碼能力有關聯。

Shiotsu 與 Weir (2007) 對詞彙和語法知識對第二語言的閱讀理解進行討論，並且探討語法和詞彙知識在預測文本閱讀理解方面的相對重要性。該研究結果發現，句法知識相對於詞彙知識而言，在預測閱讀能力方面的表現更為顯著。

Diana 等人 (2011) 的研究中，句法複雜性對於文本閱讀理解的影響不顯著，其結果原因可能是在該研究的測驗中，不同程度學生的認知負荷可能會抵消語法上簡單和複雜文本之間的差異，在本研究中，句法簡單的文本僅要求學生進行句子之間的邏輯聯繫，而句法複雜的文本要求學生在短時間內分析大量的複句結構以理解內文，句法難易度的過度落差導致測驗結果無法分析句法對文本閱讀的影響。即便在該研究中句法複雜性對於文本閱讀理解的影響不顯著，但仍然沒否定句法複雜性對文本的影響，還將句法複雜性列入為來重要的研究方向。

而在中文為二語的句法複雜性研究中，Shen 與 Tao (2011) 針對中文詞彙廣度和句法知識在閱讀理解預測中的相對關係進行分析，研究結果顯示，學習者在閱讀理解測試中的表現與他們的詞彙廣度和句法知識之間存在正相關；此外，句法知識的預測能力勝過詞彙廣度，並能更精確預測閱讀理解能力。文中進一步提到，無論學生獲得了多少個單詞，若他們無法理解句子的句法結構，他們仍然無法理解閱讀文本的內容。

宋曜廷等人 (2013) 在探討中文文本的可讀性研究中，句法難度的指標選取了 7 項，包含了句子中涵蓋的實詞數、否定詞、複雜句數、句平均詞數、單句數比率、名詞片語修飾語數以及名詞片語比率，其中詞類雖然是重要的語法複雜性，但同時也可以是詞彙複雜性的參考依據。

Lu 等人 (2019) 分析科學寫作與科學影響之間的關係，該研究中句法難度的指標以句子長度和句子複雜性作為難度評比，研究結果發現，跨學科觀察到各種修飾語結構，平均句子長度為 24.9 個單詞，標準差為 11.6，通常 126.7 個單詞組成一個段落。

此外，還有許多對於中文語法複雜性和句子難度的分析集中在單一語法點上。如張莉萍 (2005) 根據教學語法理論提出「比」字句的教學排序建議，



以中文句結構樹資料庫整理帶有比較功能的句式，其研究結果發現「引介比較句」是最常使用的表達方式，再以中央研究院平衡語料庫查詢「比」字句並列出教學排序。肖奚強等人（2009）從有中文特色的 20 個常用句，從句法和語義的角度，進行必要的下位分類及下位類之間的學習難度和分級排序。

陳琚（2010）則以韓國學習者書面與口語作為語料來源，考察「比」字句各種句型的習得難度和順序，根據五個教材大綱歸納挑選出 20 個「比」字句，再根據調查實際情況增加 2 個比較句式。Hong（2018）以華語寫作語料庫為本，分析五種不同類型的中文關聯連接詞語料，以了解關聯連接詞之間不同的句法結構以及學習者偏誤情形，並提供關聯連接詞於寫作教學的教學參考與建議。王昱勻（2021）則以 Huang（1999）和 Huang 與 Liu（2014）的分類標準與分析結構為基礎，將被動句分為長被動句和短被動句、直接被動句和間接被動句，以及新型被動句等三類進行探討，並提供教學排序，認為一開始應先教結構最簡單的短被動句，再教授結構中帶有施事者的長被動句，接著是直接被動句與間接被動句，至於第三類的新型被動句可依據學習者學習狀況當作補充教材。

綜上所述，中文句子難度的發展從過去探討理解力到指標的建立以及驗證，不外乎都會討論句子難度的評比，並且討論詞彙以及句法對句子的影響，在教學上，亦可做為不同程度之中文作為二語教學排序的參考指標，然而各研究之間對於詞彙複雜性及語法複雜性的評比目前尚未取得共識，指標的選定也都是根據各研究的成果來確立，所以建立具客觀性、科學性、規範性的語法難度指標是有其重要性與價值性。

#### 4. 研究方法

本研究為了探討影響以中文作為二語學習者對於中文文本理解的句子難度，以語料庫語言學為本建立具系統性及客觀性的中文句子難度指標，採用「中文語法數位平台<sup>1</sup>」的句子種類、句子架構的語法點為指標建立的主要來源，再以「華語八千詞表」與「文本可讀性指標自動化分析系統 3.0<sup>2</sup>」加以驗證。為增加指標之客觀性，本研究採用「國際中文教育中文水平等級標準」進行比對，以期能使指標更加客觀與準確。以下將針對「中文語法數位平台」

<sup>1</sup> 中文語法數位平台：<http://140.122.63.128/SyntaxSystem/>

<sup>2</sup> 文本可讀性指標自動化分析系統 3.0：  
<http://www.chinesereadability.net/CRIE/?LANG=CHT>

「華語八千詞表」、「文本可讀性指標自動化分析系統 3.0」以及「國際中文教育中文水平等級標準」分別進行說明。

#### 4.1 中文語法數位平台

「中文語法數位平台」是以階層式建立語法點的語法庫數位平台（洪嘉馥 2021），其首頁呈現如下圖 1，以整合「教與學」為主要目標，根據頻率與實用性篩選語法點，並分為「句子種類」、「句子架構」、「句子主要成分」等三大類型，每一大類可再細分為語法點、語法點次層、語法點細類。



圖 1：中文語法數位平台首頁

本研究為了解各項語法點的詞彙複雜性和語法複雜性，提供客觀且符合學習者程度的語法指標，採用「中文語法數位平台」的四大頻率塊作為參考依據。「中文語法數位平台」的四大頻率塊針對每一項語法點，以母語者使用頻率、二語學習者的使用偏誤頻率做為象限，區分成四大頻率塊，由第一象限到第四象限分別是「常出現常錯誤」、「常出現不常錯誤」、「不常出現不常錯誤」及「不常出現常錯誤」。「常出現不常錯誤」即指出現頻率高且偏誤率低的語法點；「不常出現不常錯誤」是指出現頻率低且偏誤率低的語法點；「常出現常錯誤」為出現頻率高且偏誤較高的語法點；「不常出現常錯誤」則為出現頻率低且偏誤率高的語法點，呈現如下圖 2 所示，並以句子架構中的複句語法點「但（是）」為例。語法點「但（是）」是建設在句子架構下的複句，

因其用來表示轉折語氣，所以再細歸於轉折關係中，而在頻率塊中，其屬於「常出現常錯誤」，可知在語法複雜性中，語法點「但（是）」屬於較簡單的語法點。



圖 2：中文語法數位平台語法點「但（是）」

#### 4.2 華語八千詞表

本研究在句子難度指標的建立上，以詞彙複雜性和語法複雜性做為計量指標，並考察中文的語義特徵。關於詞彙複雜性，除了參考過去文獻以及「中文語法數位平台」，也以「華語八千詞表」作為依據，考察各級詞彙的難度標準制定。

《華語八千詞表》是由國家華語測驗推動工作委員會（以下簡稱華測會）對中文學習者所製定兼具信度和效度的華語能力測驗標準，是與華語文能力測驗相應的學習核心詞表。故本研究將《華語八千詞表》與「中文語法數位平台」進行比對，作為詞彙複雜性之分類依據。

#### 4.3 文本可讀性指標自動化分析系統 3.0

文本可讀性指標自動化分析系統（Chinese Readability Index Explorer，

CRIE) 能夠自動分析文本多項特徵，以體現閱讀材料的可讀性 (readability)，亦即了解讀者理解與掌握資訊的程度。當中，系統能看出文本中華語詞彙難度，將有效地確認「中文語法數位平台」當中「句子架構」以及「句子成分」之細項難度，因此，在此次研究中也將採用該系統收錄的母語文本語料與二語語料作為句子難度指標的驗證依據，其首頁如圖 3。



圖 3：文本可讀性指標自動化分析系統 CRIE 3.0 首頁

#### 4.4 國際中文教育中文水平等級標準

由於《國際中文教育中文水平等級標準》(以下簡稱標準)是眾多難度配置的參考標準，清楚制定了音節、漢字、詞彙、語法的具體量化指標，並將難度分為初等、中等、高等，三類之下又細分一至九級的三等九級漢語水準等級標準 (HSK)，並且作為本研究驗證語法點難度的參考基礎。

#### 5. 中文句子難度指標

本研究為了確認中文文本的句子難度，故擷取「中文語法數位平台」中

針對句子結構劃分的「句子架構」類及「句子主要成分」類之細項為依據，透過四個頻率塊的分析結果，將平台無法確認難度且具語法功能的詞彙對比「華語八千詞表」以取得難度，並以「CRIE」確立影響句子難度的詞彙指標與語法指標。指標的判斷標準為詞彙與語法的母語環境「出現頻率」以及學習者的「出現錯誤頻率」。出現頻率高的語法點應較出現頻率低的語法點早出現，出錯率的高低則代表該語法點的易掌握程度，本研究透過 CRIE 中的母語語料分析所有語法點的「出現頻率」，並對比「中文語法數位平台」四個頻率塊的分類結果；同時亦採用 CRIE 中的二語語料分析所有語法點的「出現錯誤頻率」，並對比「中文語法數位平台」四個頻率塊的分類結果；最後依據四個頻率塊將難度由簡單至難的排序為，「常出現不常錯誤」、「不常出現不常錯誤」、「常出現常錯誤」及「不常出現常錯誤」。以下根據四大頻率塊之詞彙與語法點排序中，出現頻率與偏誤率皆高的程度副詞「很」以及「最」為例（請參附錄一），CRIE 的對比結果呈現如下圖 4 與圖 5。

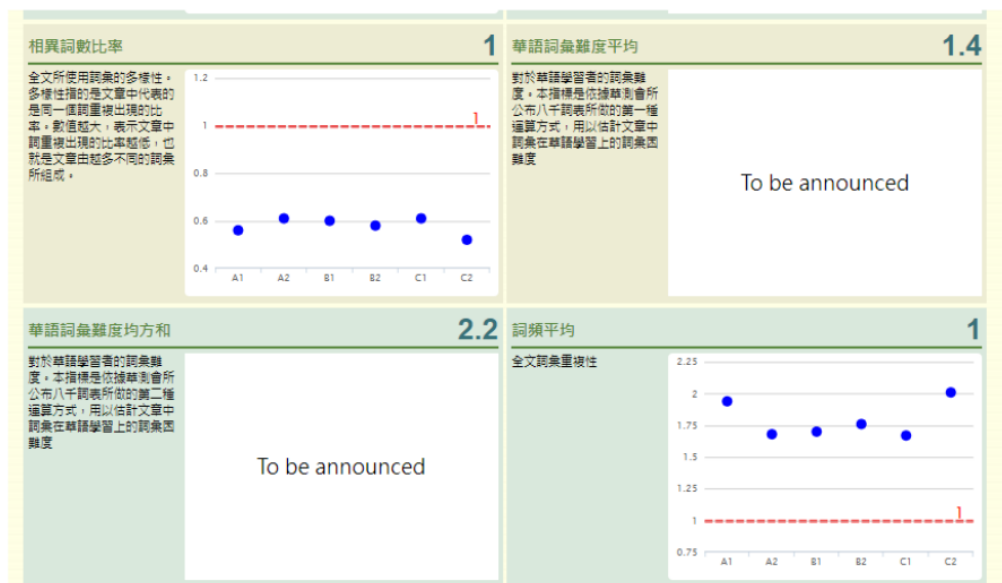


圖 4：文本可讀性指標自動化分析系統驗證程度副詞「很」之難度

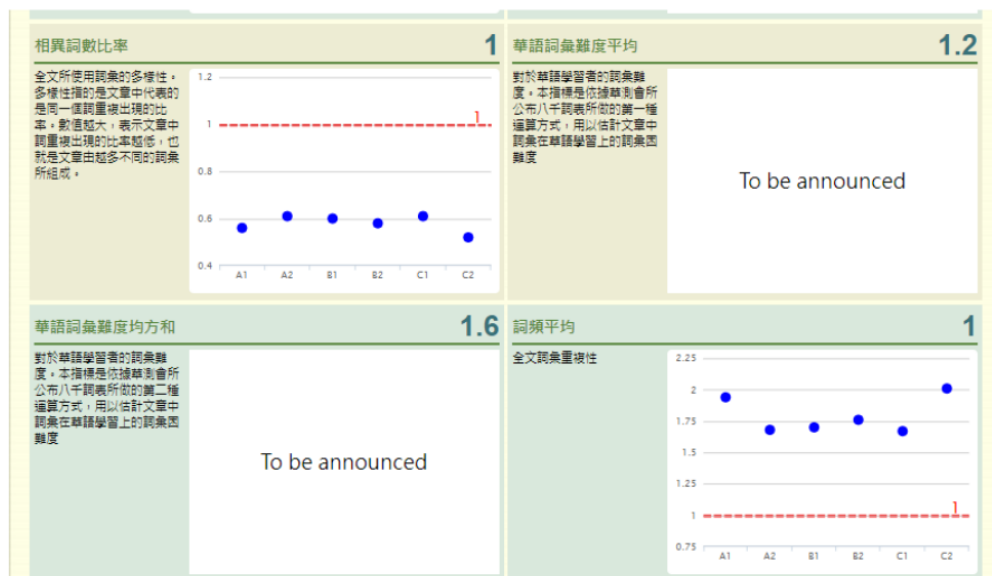


圖 5：文本可讀性指標自動化分析系統驗證程度副詞「最」之難度

根據四大頻率塊之詞彙與語法點的排序（請參附錄一）判斷，對二語學習者而言，「很」相較於「最」是較為簡單易學的；而透過 CRIE 進行應證，呈現如下圖 4 與圖 5，亦證實「很」相較於「最」，詞彙難度平均較低，對學習者而言屬於較為容易的程度副詞。

「句子架構」的各項語法點在四頻率塊的比例分析結果中（請參附錄二），多數的語法類型都有顯著的頻率傾向，如「單句-對等關係」，這一類型語法點在「常出現不常錯誤」的比例為 50%，「常出現常錯誤」為 25%，「不常出現不常錯誤」為 25%，可判斷「單句—對等關係」屬於「常出現不常錯誤」難度低的語法類型；此外，與「單句的對等關係」在「常出現不常錯誤」比例同樣為 50%的「偏正複句—目的複句」，由於在「常出現常錯誤」的表現為 50%，比「單句—對等關係」的 25%還要高，因此可推斷「單句—對等關係」類型的語法點比「偏正複句—目的關係」的類型簡單。


「句子成分」的各項語法點在四頻率塊的比例分析結果中（請參附錄三），多數的語法類型也都有明確的頻率傾向，不過比「句子架構」的結果表現更為顯著，如「動詞『有』」在「常出現不常錯誤」的比例為 100%；「數量副詞」在「常出現常錯誤」的比例亦為 100%。

由於部分指標的難度較難直接確認與判斷，例如：「句子架構」下的「列



舉關係」、「承接關係」和「可能補語」，以及「句子成分」當中的「量詞」，本研究為了補足這兩種類型語法，因此，本研究透過「華語八千詞表」及「CRIE」進行人工排查，分析出相對應之語法點順序，其排序如下表 1 所示。


表 1：「句子架構」類語法點難度順序

「句子架構」類語法點	
<p>難度低</p>  <p>難度高</p>	單句—對等關係
	單句—列舉關係
	聯合複句—遞進關係
	聯合複句—並列關係
	偏正複句—因果關係
	偏正複句—目的關係
	聯合複句—選擇關係
	動補結構—結果補語
	偏正複句—假設關係
	偏正複句—轉折關係
	偏正複句—條件關係
	聯合複句—承接關係
	動補結構—數量補語
	動補結構—趨向補語
	偏正複句—讓步關係
	動補結構—可能補語
	動補結構—情態補語
	動賓複句
	複雜句

由上表 1 可見，複句的語法類型中，即便是同一類型之語法，難易程度仍然有所不同，但可以確定句子架構會隨著複雜程度提高而增加難度，也因此，根據分析結果的難度排序，依序為單句、複句、複雜句。至於針對句子

成分類語法點的排序也是相似的，成分類型相同的語法點之間難度也有明顯的不同，其排序如下表 2 所示。

表 2：「句子成分」類之語法點難度順序

「句子成分」類語法點	
<p>難度低</p>  <p>難度高</p>	普通名詞
	專有名詞
	地方名詞
	代名詞
	時間名詞
	及物動詞
	不及物動詞
	動詞「是」
	動詞「有」
	名量詞
	約量詞
	能願動詞
	借詞
	數量副詞
	否定詞
	結構語助詞
	頻率副詞
	程度副詞
	範圍副詞
	時間副詞
	語氣副詞
	評價副詞
	動量詞



	時貌
	語氣助詞
	嘆詞

根據上表 2 的結果，句子成分的難度排序中，名詞類佔據前幾名，亦是學習者最先接觸的類型，而隨著學習者所能組織的句子長度及複雜度，而後才會學習動詞、副詞、時貌等句子成分。

為確立各類語法點難度的客觀性，本研究參閱「標準」中語法點的等級配置，對比 CEFR 與 HSK 中各語法點之難度等級，兩者皆以二語學習者的學習歷程為主要考量，但對比後發現兩者的難度分佈有幾處不同之處，考量學習者首次接觸某一語法類型時，所需的先備知識與中文能力存在階段性的差異，因此以各語法類型首次出現的語法點之難度進行整理，依據使用頻率、偏誤率，以及首次出現的語法點之難度等級進行劃分成 4 個等級，「句子架構」類語法點與「句子成分」類語法點的分布結果，分別呈現於下表 3 及表 4。

表 3：「句子架構」類語法點之難度等級分佈

「句子架構」	
難度等級	語法點類型
1	單句—對等關係
2	單句—列舉關係
	聯合複句—遞進關係
	聯合複句—並列關係
	偏正複句—因果關係
	聯合複句—選擇關係
	動補結構—結果補語
	偏正複句—轉折關係
	偏正複句—假設關係
	偏正複句—條件關係
	聯合複句—承接關係

	動補結構—數量補語
	動補結構—趨向補語
	偏正複句—讓步關係
3	偏正複句—目的關係
	動補結構—可能補語
	動補結構—情態補語
	動賓複句
4	複雜句

由上表 3 所示，句子架構的難度等級主要依循句子的複雜程度分為四個等級，針對複句的部分，考量語法點的使用頻率、偏誤率，以及所使用的詞彙難度，故又細分成兩個難度等級。而句子成分類語法點首次出現的難度都較為簡單，僅有兩個難度等級，結果呈現如下表 4。

表 4：「句子成分」類語法點之難度等級分佈

「句子成分」	
難度等級	語法點類型
1	普通名詞
	專有名詞
	地方名詞
	代名詞
	及物動詞
	不及物動詞
	動詞「是」
	動詞「有」
	名量詞
	約量詞
	借詞

	否定詞
	結構語助詞
	頻率副詞
	時間副詞
	程度副詞
	範圍副詞
	語氣助詞
	時貌
2	時間名詞
	能願動詞
	數量副詞
	語氣副詞
	評價副詞
	動量詞
	嘆詞

依據上表 4，句子成分常以詞彙形式出現，對比句子架構普遍難度不高，在學習者完全習得某一語法點的應用前，便有機會接觸句子成分中的任一語法點。不過，無論是句子架構或是句子成分的難度等及分布，可見部分語法難度本身不高，如「目的複句」和「時間名詞」，但對學習者而言，須在具備一定的中文能力以及語法知識的前提下，才得以順利習得，因此此兩類型語法整體難度較高，出現的時間亦較晚，需待學習者的語言能力提升後，才能掌握並加以應用。

本研究以中文語法數位平台的四大頻率塊之功能，將各類型之語法點進行使用頻率與偏誤率的量化統計，考量指標之客觀性與完整性，以「華語八千詞表」及「CRIE」進行比對，再以「標準」進行驗證，建立中文句子難度指標。中文句子難度指標依據語法類型首次出現之語法點的難度等級、使用頻率，以及偏誤率可劃分為四個等級，可作為教學上教材設計的參考依據，也能作為檢視中文為二語學習者學習進度的指標。

## 6. 結論

句子是閱讀理解的基礎，而組成句子的成分包含詞彙與語法，兩者的複雜性是華語教學所關注的議題之一，不同難度的教材，內容所選用的詞彙與語法皆會影響學習者的學習與表現，也因此建立完善且客觀的難度指標至關重要。

中文句子難度一般所探討的面向主要分為詞彙複雜性與語法複雜性，且常僅關注於單一面向，或以單一類型進行分析，有鑑於此，本研究為了提供客觀且符合學習者程度的語法點難度，從文本的角度出發，探討文本中句子之間的難度差異，再以語料庫語言學為本，採用 CRIE 中的母語語料與二語語料，驗證中文語法數位平台中各類型語法之語法點之使用頻率與偏誤率，並探究詞彙複雜性與語法複雜性對中文句子難度的影響，建立中文句子難度之指標，最後採用「標準」加以驗證，發現「華語八千詞表」和「標準」的難度分級因為語料來源不同，在難度分級上出現差異，故而以語法類型首次出現之語法點的難度為該語法類型之難度，並進行難度分級，建立具系統性及客觀性的中文語法點難度排序。

本研究的中文句子難度指標可作為教學設計的參考依據，根據學習者的學習程度調整語法學習，或準備更為合適的學習內容，同時，教師亦可透過此指標追蹤學習者的學習狀況以提供的學習材料，進而提升學習者的句子理解能力與閱讀能力。此外，本研究目前的成果是以文本角度切入的整體發展指標，透過母語者的使用頻率與二語學習者的偏誤率取得高度客觀性，然本研究仰賴量化分析的結果，未考慮語言習得的歷程與語料來源的多元性，因此未來尚可根據中文為二語之學習者的習得歷程、教學教材等層面深入探討語法點之間的具體差異，或者，基於語料庫的語料來源很多元，應考量不同領域所使用之語法點的難度與使用情形不同，加入不同領域的語料加以分析，以完善此語法點難度指標。

## 引用文獻

- Ai, Haiyang. 2022. Automating Lexical Complexity Measurement in Chinese with WeCLECA. *International Journal of Asian Language Processing* 32.1: 1-18.
- Bowey, J. A. 1986. Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology* 4.2: 282-299.

- Covington, M. A., He, C., Brown, C., Naci, L., and Brown, J. 2006. How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level Scale.
- Diana J. Arya, Elfrieda H. Hiebert, P. David Pearson. 2011. The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education* 4.1: 107-125.
- Eisenberg, S. 2002. Interpretation of relative clauses by young children: another look. *Journal of Child Language* 29: 177-188.
- Gordon P., Hendrick R., and Johnson M. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51: 97-114.
- Gunning, R. 1952. The technique of clear writing. New York: McGraw-Hill.
- Hong, J. F. 2018. Chinese Conjunctions in Second Language Learners' Written Texts. *Lecture Notes in Artificial Intelligence (LNAI)*, 10709: 510-522.
- Hong, Jia-Fei, Chun-Yi Peng, Hao-Chiang Tseng, and Yao-Ting Sung. 2020. Linguistic Feature Analysis of CEFR Labeling Reliability and Validity in Language Textbooks. *Journal of Technology and Chinese Language Teaching* 11.1: 57-83.
- Hong, J. F., Sung, Y. T., Tseng, H. C., Chang, K. E., Chen, J.L. 2016. A multilevel analysis of the linguistic features affecting Chinese text readability. *Taiwan Journal of Chinese as a Second Language* 13: 95-126.
- Huang, C.-T. James. 1999. Chinese passives incomparative perspective. *Tsing Hua Journal of Chinese Studies* 29.4: 423-509.
- Huang, C.-T. James, and Na Liu. 2014. A new passive form in Mandarin: Its syntax and implications. *International Journal of Chinese Linguistics* 1: 1-34.
- Kenneth, I. F., and Susan, M. C. 1973 Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12: 627-635.
- Kidd E. 2003. Relative clause comprehension revisited: commentary on Eisenberg (2002). *Journat of Child Language* 30: 671-679.
- Klare, G. R. 2000. The measurement of Readability: Useful information for communicators. *ACM Journal of Computer Documentation* 24: 107-121.
- Laufer, B., and Nation, P. 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16: 307-322.

- Lu, C., Bu, Y., Dong, X., Wang, J., Ding, Y., Larivière, V., Sugimoto, C. R., Paul, L., and Zhang, C. 2019. Analyzing linguistic complexity and scientific impact. *Journal of Informetrics* 13.3: 817-829.
- Lu, X. F. 2008. Automatic Measurement of Syntactic Complexity Using the Revised Developmental Level Scale. Paper presented at the FLAIRS Conference, 153-158, May, 2008. Florida, USA.
- Lu, X. F. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15.4: 474-496.
- Lu, X. F., and Ai, H. Y. 2015. Syntactic Complexity in College-Level English Writing: Differences among Writers with Diverse L1 Backgrounds. *Journal of Second Language Writing* 29: 16-27.
- McLaughlin, G. H. 1969. SMOG grading -- A new readability formula. *Journal of Reading* 22: 639-646.
- Rosenberg, S., and Abbeduto, L. 1987. Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults. *Applied Psycholinguistics* 8.1: 19-32.
- Shen, Y. L., and Tao, W. 2011. The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of Reading Comprehension Test Performance. *Chinese Journal of Applied Linguistics* 34.3: 113-126.
- Shiotsu, T., and Weir, C. J. 2007. The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing* 24.1: 99-128.
- Silberstein, S. 1987. Let's take another look at reading: Twenty-five years of reading instruction. *English Teaching Forum* 25: 28-35.
- Spache, G. 1953. A new readability formula for primary-grade reading materials. *Elementary school journal* 53: 410-413.
- Thoma, A. G., and Gilbert, K. K. 1967. Influence of Syntactic Errors on Sentence Recognition. *Journal of Verbal Learning and Verbal Behavior* 6: 692-698.
- Tsao, F. F. 1996. ON VERB CLASSIFICATION IN CHINESE. *Journal of Chinese Linguistics* 24.1: 138-191.
- Ye, W. X., and Qiu L. K. 2008. Hanyu fuheci lijie nanyidu de jisuan [〈漢語複合詞理解難易度的計算〉Computing the degree of readability in understanding

- Chinese compounds.]. *Language and Linguistics* 9.2: 435-447.
- 于東、吳思遠、耿朝陽、唐玉玲。2020。〈基於眾包標注的語文教材句子難易度評估研究〉，《中文信息學報》，第2期，1-11。[Yu, Dong, Wu, Si-Yuan, Geng, Zhao-Yang and Tang, Yu-Ling. 2020. Assessing Sentence Difficulty in Chinese Textbooks Based on Crowdsourcing. *Journal of Chinese Information Processing*, 2: 1-11.]
- 王昱勻。2021。〈華語被字句的分類及其教學建議〉，《華文世界》，第127期，76-89。[Wang, Yu-Jun. 2021. Huáyǔ bèizìjù de fēnlèi jí qí jiāoxué jiànyì. *The World of Chinese Language*, 127: 76-89.]
- 王進、周慧、羅國峰、顧翔。2017。〈基于自然語言處理的圖書閱讀難度自動分級研究〉，《計算機時代》，第8期，1-5。[Wang, Jin, Zhou, Hui, Luo, Guo-Feng and Gu, Xiang. 2017. Jī yú zìrán yǔyán chǔlǐ de túshū yuèdú nándù zìdòng fēnjí yánjiù. *Computer Era*, 8: 1-5.]
- 吳思遠、于東、江新。2020。〈漢語文本可讀性特征體系構建和效度驗證〉，《世界漢語教學》，第1期，81-97。[Wu, Si-Yuan, Yu, Dong and Jiang, Xin. 2020. Development of Linguistic Features System for Chinese Text Readability Assessment and Its Validity Verification. *Chinese Teaching in the World*, 1: 81-97.]
- 吳默。2021。〈基於教學層面之話語標記「並」的語用功能初探〉，《華語學刊》，第30期，19-41。[Wu, Mo. 2021. An Investigation of the Pragmatic Function of the Discourse Marker Bing Based on the Pedagogical Level. *TCSL Forum*, 30: 19-41.]
- 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》（簡體書）。中國：高等教育出版社。[Xiao, Xi-Qiang. 2009. Wàiguó xuéshēng hànyǔ jùshì xuéxí nándù jí fēnjí páixù yánjiù. Beijing: Higher Education Press.]
- 宋曜廷、陳茹玲、李宜憲、查日穌、曾厚強、林維駿、張道行、張國恩。2013。〈中文文本可讀性探討：指標選取、模型建立與效度驗證〉，《中華心理學刊》，第55卷第1期，75-106。[Sung, Yao-Ying, Chen, Ju-Ling, Lee, Yi-Shian, Cha, Jih-Ho, Tseng, Hou-Chiang, Lin, Wei-Chun, Chang, Tao-Hsing and Chang, Kuo-En. 2013. Investigating Chinese Text Readability: Linguistic Features, Modeling, and Validation. *Chinese Journal of Psychology*, 55.1: 75-

106.]

竺家寧。1999。《漢語詞彙學》。台北市：五南。[Chu, Chia-Ning. 1999. *Hanyu Lexicology*. Taipei: Wu-Nan Book Inc.]

洪嘉馥。2021。〈以語料庫為本建置中文語法數位平台及其輔助華語教學〉，《華語文教學研究》，第 18 卷第 1 期，59-87。[Hong, Jia-Fei. 2021. A Corpus-based Study of Development of the Digital Platform of Chinese Grammar Bank and its Assistant in CSL. *Journal of Chinese Language Teaching*, 18.1: 59-87.]

唐玉玲、于東。2020。〈結合深度學習和語言難度特徵的句子可讀性計算方法〉，發表於「第十九屆中國計算語言學大會論文集」，2020 年 10 月 30 日-11 月 1 日。中國海南。[Tang, Yu-Ling, Zhang, Yu, Dong. 2020. Combination of Deep Learning and Language Difficulty Feature for Sentence Readability Metric. Paper presented at the 19<sup>th</sup> China National Conference on Computational Linguistics, October 2020. Hainan.]

孫雪芹。2023。〈華語文能力基準與語義場關聯詞查詢系統應用〉，《華文世界》，第 132 期，219-224。[Soon, Suet-Ching. 2023. Huáyǔwén Nénglǐ Jīzhǔn yǔ Yǔyìchǎng Guānliáncí Cháxún Xitǒng Yīngyòng. *The World of Chinese Language*, 132: 219-224.]

張莉萍。2005。〈淺談教學語法——從比較句談起〉，發表於「第八屆漢語教學討論會」，2005 年 7 月 23-25 日。中國北京。[Chang, Li-Ping. 2005. Qiǎntán Jiāoxué Yǔfǎ — Cóng Bǐjiào jù Tánqǐ. Paper presented at the 8<sup>th</sup> Hànyǔ Jiāoxué Tǎolùnhuì, July 2005. Beijing.]

張莉萍。2017。〈華語學習者關係子句的習得考察——以語料庫為本的研究〉，《華語文教學研究》，第 14 卷第 1 期，47-80。[Chang, Li-Ping. 2017. The acquisition of relative clauses in L2 Chinese: a corpus-based study. *Journal of Chinese Language Teaching*, 14:1: 47-80.]

陳司恩。2019。〈近義詞辨析：「改正」、「糾正」、「矯正」〉，《臺大華語文教學研究》，第 7 期，69-90。[Chen, Si-En. 2019. The Distinction Between Dynonyms: gǎi zhèng, jiū zhèng and jiǎo zhèng. *NTU Working Papers in Chinese Language Teaching*, 7: 69-90.]

陳旻祺。2013。〈華語語氣副詞「幸虧」、「幸好」、「好在」與「多虧」的比較——以語料庫為本〉，《臺大華語文教學研究》，第 1 期，39-63。[Chen, M



- in- Qi. 2013. Huáyǔ Yǔqì Fùcí: Xíngkuī、Xíng hǎo、Hǎozài yǔ Duō kuī de bǐjiào — Yǐ Yǔliàokù Wéi Běn. *NTU Working Papers in Chinese Language Teaching*, 1: 39-63.]
- 陳琚。2010。〈比較句語法項目的習得難度考察〉，《華南師範大學學報（社會科學版）》，第3期，46-52。[Chen, Jun. 2010. Bǐjiàojù Yǔfǎ Xiàngmù de Xidé Nándù Kǎochá. *Journal of South China Normal University : Social Science Edition*, 3: 46-52.]
- 楊惠媚、陳浩然、潘依婷。2015。〈分析華語口語語料庫高頻詞之特點並對 TOCFL 詞表提出建議〉，《華語文教學研究》，第12卷第1期，1-44。[Yang, Hui-Mei, Chen, Howard Hao-Jan and Pan, I-Ting. 2015. Analyzing the Features of the High-Frequency Words on Chinese Spoken Corpus and Offering the Word-recruiting Suggestion to TOCFL Wordlist. *Journal of Chinese Language Teaching*, 12.1: 1-44.]
- 鄭錦全。2005。〈詞彙語意與句子閱讀難易度計量〉，《第六屆漢語詞彙語意學研討會論文集》，261-265。[Cheng, Chin-Chuan. 2005. Cíhuì Yǔyì yǔ Jùzi Yuèdú Nányìdù Jiliáng *The Sixth Chinese Lexical Semantics Workshop*, 261-265.]
- 蔡美智。2010。〈華語近義詞辨識難易度與學習策略初探〉，《臺灣華語教學研究》，第1期，57-79。[Tsai, Mei-Chih. 2010. Preliminary Study on Strategies for Learning Near-Synonyms in TCSL. *Taiwan Journal of Chinese as a Second Language*. 1: 57-79.]
- 龐成。2016。〈漢語句子難易度影響因素分析〉，《語文學刊》，第1期，18-19。[Pang, Cheng. 2016. Hànyǔ Jùzi Nányìdù Yǐngxiǎng Yīnsù Fēnxī. *Journal of Language and Literature Studies*, 1: 18-19.]

[審查：2024.5.28 修改：2024.9.9 接受：2022.9.16]

洪嘉馥

Jia-Fei HONG

10610 臺北市和平東路一段 162 號 國立臺灣師範大學華語文教學系

Department of Chinese as a Second Language

National Taiwan Normal University

華語文教學研究

No.162, Sec. 1, Heping E. Rd., Taipei City 10610, Taiwan  
jiafeihong@ntnu.edu.tw

陳家妮

Jia-Ni CHEN

10610 臺北市和平東路一段 162 號 國立臺灣師範大學華語文教學系  
Department of Chinese as a Second Language  
National Taiwan Normal University  
No.162, Sec. 1, Heping E. Rd., Taipei City 10610, Taiwan  
jiani@ntnu.edu.tw

## 附錄一 按照四大頻率塊之詞彙與語法點排序

程度副詞	二語	母語	結果
更	21	398359	常出現不常錯誤
很	343	260190	常出現常錯誤
最	61	666990	常出現常錯誤
非常	23	96980	不常出現不常錯誤
真	16	25884	不常出現不常錯誤
比較	15	28546	不常出現不常錯誤
好	74	8050	不常出現常錯誤
稍微	搜尋後無結果	3242	無法判斷

## 附錄二 「句子架構」細項之四頻率塊比例分析

句子結構	語法類型		常出現不常錯誤		常出現常錯誤		不常出現不常錯誤		不常出現常錯誤		無法判斷
單句	對等		2	50%	1	25%	1	25%	0	0%	0
	列舉		無建置								
複句	聯合複句	遞進關係	2	25%	0	0%	5	63%	1	13%	3
		並列關係	3	33%	1	11%	5	56%	0	0%	1
		選擇關係	0	0%	0	0%	4	100%	0	0%	2
		承接關係	無建置								
	偏正複句	目的複句	1	50%	1	50%	0	0%	0	0%	3
		因果關係	3	43%	0	0%	3	43%	1	14%	3
		轉折關係	2	22%	1	11%	6	67%	0	0%	5
		假設關係	1	33%	0	0%	2	67%	0	0%	11
		條件關係	0	0%	0	0%	8	100%	0	0%	3
		讓步關係	0	0%	0	0%	4	100%	0	0%	5
	動補結構	可能補語	無建置								
		結果補語	0	0%	1	50%	1	50%	0	0%	2
		數量補語	0	0%	0	0%	0	0%	0	0%	3
		趨向補語	0	0%	0	0%	6	100%	0	0%	0
		情態補語	0	0%	0	0%	2	100%	0	0%	1
	動賓複句		0	0%	0	0%	0	0%	0	0%	6
複雜句	關係子句		0	0%	0	0%	0	0%	0	0%	4

### 附錄三 「句子成分」細項之四頻率塊比例分析

句子成分	語法類型	常出現不常錯誤		常出現常錯誤		不常出現不常錯誤		不常出現常錯誤		無法判斷
名詞	1.普通名詞	0	0%	0	0%	0	0%	0	0%	1
	4.地方名詞	0	0%	0	0%	0	0%	0	0%	1
	2.專有名詞	0	0%	0	0%	0	0%	0	0%	1
	3.時間名詞	0	0%	0	0%	0	0%	0	0%	1
	5.代名詞	0	0%	0	0%	0	0%	0	0%	9
動詞	6.及物動詞	0	0%	0	0%	0	0%	0	0%	1
	7.不及物動詞	0	0%	0	0%	0	0%	0	0%	1
	8.能願動詞	1	10%	2	20%	4	40%	3	30%	2
	9.動詞「是」	0	0%	0	0%	0	0%	0	0%	4
	10.動詞「有」	1	100%	0	0%	0	0%	0	0%	3
量詞	11.名量詞	0	0%	0	0%	0	0%	0	0%	1
	12.動量詞	0	0%	0	0%	0	0%	0	0%	1
	13.約量詞	0	0%	0	0%	0	0%	0	0%	1
副詞	14.頻率副詞	1	20%	0	0%	4	80%	0	0%	0
	15.數量副詞	0	0%	1	100%	0	0%	0	0%	2
	16.時間副詞	0	0%	2	13%	12	80%	1	7%	8
	17.語氣副詞	0	0%	0	0%	14	100%	0	0%	10
	18.範圍副詞	0	0%	1	33%	2	67%	0	0%	0
	19.評價副詞	0	0%	0	0%	0	0%	0	0%	1
	20.程度副詞	1	0%	2	29%	3	43%	1	14%	0
功能詞	21.否定詞	1	20%	1	20%	3	60%	0	0%	2
	22.介詞	6	67%	1	11%	2	22%	0	0%	0
	23.結構語助詞	1	20%	3	60%	0	0%	1	20%	0
	24.語氣助詞	0	0%	0	0%	3	100%	0	0%	11
	25.嘆詞	0	0%	0	0%	1	100%	0	0%	6
	26.時貌	1	17%	1	17%	2	33%	2	33%	9

## **A Study on the Difficulty of Chinese Sentences Affecting the Readability of Chinese Texts**

**Jia-Fei HONG**  
**Department of Chinese as a**  
**Second Language**  
**National Taiwan Normal**  
**University**

**Jia-Ni CHEN**  
**Department of Chinese as a**  
**Second Language**  
**National Taiwan Normal**  
**University**

### **Abstract**

The importance of reading comprehension has been increasingly recognized in recent years, with its wide-ranging impact. Whether in the acquisition of Chinese as first language or as second language, reading comprehension is considered as one of crucial skills. The influence of reading comprehension spans from vocabulary and grammar to discourse, all of which involve the difficulty and complexity of sentences (Hong et al. 2020; Tang and Yu 2020). Therefore, in teaching Chinese as a second language, the vocabulary and grammar that make up sentences are often used as criteria for accessing reading proficiency (Silberstein 1987). However, existing studies on the difficulty of Chinese sentences often focus on single indicators. Some studies only discuss the complexity of lexical features (Ye and Qiu 2008; Tsai 2010; Chen 2013; Chen 2019), while others focus on the diversity of specific grammatical structures (Chang 2005 ; Xiao et al. 2009; Chen 2010; Wang 2021). Some studies even lack integrated and objective quantitative standards (Cheng 2005; Ye and Qiu 2008; Pang 2016; Wu et al. 2020).

Thus, this study, grounded in corpus linguistics, investigates the influence of lexical and syntactic complexity on the difficulty of Chinese sentences through the Digital Platform of Chinese Grammar. It quantitatively ranks the difficulty of vocabulary and grammar, and validates these rankings using the 8000 Vocabulary List, the Chinese Readability Index Explorer 3.0, and the Chinese Proficiency Grading Standards for International Chinese Language Education. By comparing

the difficulty of various grammatical structures, this study establishes a systematic and objective ranking of Chinese grammar difficulty. This ranking aims to assist in the teaching of Chinese sentence, enhancing the sentence comprehension abilities and overall reading proficiency of Chinese as a second language learners.

**Keywords:** lexical complexity, reading comprehension, sentence difficulty, syntactic complexity, text readability