

國立臺灣大學理學院心理學研究所



博士論文

Graduate Institute of Psychology

College of Science

National Taiwan University

Doctoral Dissertation

診斷式的適讀性評估系統：

以小學文本探討四種模式的比較研究

Readability Diagnosis System: A Comparative Study
of Four Models on Elementary School Textbook

曾昱翔

Yu-Hsiang Tseng

指導教授：胡志偉 博士

Advisor: Chih-Wei Hue, Ph.D.

中華民國 108 年 1 月

January 2019



國立臺灣大學理學院心理學研究所

論文口試委員會審定書

曾昱翔 先生所提論文 診斷式的適讀性
評估系統：以小學文本探討四種模式的比較研究

經本委員會審議，符合博士學位標準，特此證明。

論文考試委員會

主席 許用廉

委員 許用廉

宋昭文

胡志偉

謝舒凱

呂晉晉

指導教授：胡志偉

所主任：鄭伯堃

中華民國 107 年 12 月 28 日







致謝



「這是最好的時代，也是最壞的時代。」這句話應非評價，而是在講二元，談對比；凸顯矛盾和掙扎，期待迸發的創意和擘畫可能是憧憬也可能是理想的未來。

時間是寫作論文時最大的感觸。直觀上，時間永遠是研究過程中的稀缺資源。撰寫論文時，一邊看著自己研究室桌上用鉛筆草擬的時間表，一邊掐算著自己剩餘的時間。蒐集資料時，同時寄望著電子郵件的魚雁往返，又忖量著資料蒐集時程。論文中牽涉的模型開發，我總想著有幾組模型需要嘗試，每組模型約需要幾個小時。研究的過程，就是在跟時間打交道的過程。

時間不僅體現在各種具體的作業上，也在論文寫作本身。此處的時間無疑是被抽象化的一論文陳述是根據文獻、實驗結果、實徵資料，有系統性地推出結論。所有的作品、實驗程序，縱使有時態，在書寫時還是把所有發生的事情凝結在時空中的一點，好像它們是如此地純粹必然。當它們發生之後，就像是時間流裡的結晶，在剎那間被拍下最美的瞬間，於是淬鍊成時空中的翡翠，完美地靜態呈現在幾萬字的文字晶格。

時間的痕跡也是形而上的。就如同心理學與語言學中，不斷反省著語言與認知之間的關係，我好奇的是時間和語言的關係。在符號學三角形中，語言、符號、指涉物的關係彼此密不可分，如果人們承認時間是這世界的一部份，很難想像語言與概念要如何獨立於時間之外。語言心理學在這個脈絡下是特別的，它處在語言與人類認知系統之間，它應該要有「時間感」。

我妄想探索這份「時間感」，卻花了很多時間。我在博士班的頭幾年在一系列神經造影的研究計畫中，試圖從底層視覺尋找靈感。感謝當時的計畫主持人陳建中教授和蔡紫薰醫師給我很重要的學習機會。底層視覺理論與方法的成

功，總是給我很大的希望。

在博士班期間，我有幸參與諸多重要的研究計畫，培養許多日後進行論文研究的養分。在研究團隊中，有一位最重要的老師，也是這份論文口試委員呂菁菁老師。呂老師指引我一起完成很多研究工作、更拜訪世界各地。感謝呂老師在那段時間以及最後論文階段中給予的莫大支持、協助、信任和鼓勵。

在博士班的最後幾年，我很高興參與謝舒凱老師的實驗室。感謝 LOPERs 們在最後這段時期的支持鼓勵，他們讓我感受到語言研究的活力與生命力。更感謝謝老師在計算語言學上給我的諸多指導和靈感，並擔任我的口試委員。更重要的，是讓我相信我在意的並非完全子然於世。

感謝我的口試委員宋曜廷教授，在每日繁忙公務纏身之餘，特地撥空指導我的計畫書和論文。宋老師在適讀性研究上的專業學養和觀點，幫助我有更清楚的研究方向，讓這份論文的內容更完整豐富。

感謝我的口試委員許聞廉研究員，在口試時提出很多深富經驗與智慧的評語。許研究員的意見讓我感受到語言處理的深與廣，以及各種研究取向總有不足之處。我作為後輩的研究者總是需在巨人肩上，才能找到前進的方向。

感謝我的指導教授，胡志偉老師。從大學部到博士班畢業，無數次的概念發想、研究方向、設計、文章修改、報告呈現，過程中老師的學術素養令我深感欽佩，更是感謝老師對我總是耐心且包容。如果大學時的我和現在相比，真的有多那麼一點點學術色彩，那完全是老師的功勞。

感謝我的父母，我從碩士班到博士班一路走了十多年，若不是父母的相信、尊重與支持，這一步是不可能的。也感謝我的新家人思瑩，同樣都在博士生的道路上，她總是不變的支持來源。

最後，感謝過去每個時刻的自己，以及那些存在與不存在的時間。在之中，我多看到了些什麼，也多學到了些什麼。謹此誌謝，給未來的自己。

摘要



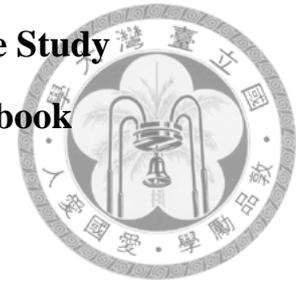
適讀性研究試圖透過分析文本的文字安排，以建立讀者閱讀時，文本易理解程度的預測模式。有研究者根據適讀性模式納入的文本屬性，以及模式建構的方式，將此類研究的發展分為三個時期。不同時期的研究在建構適讀性模式時深受「計算預測」和「認知解釋」等兩種考量的影響。在早期的傳統公式期，研究者希望從文本中找出一些屬性，然後以不同方法（如統計迴歸法）根據這些屬性，建構預測文本適讀性的模型。受到文本分析工具的限制，這時期納入適讀性模式的多是方便計算的文本表層屬性（如，文章中的難詞比例、文句的長度等）。隨著認知心理學不斷深化研究者對於人類閱讀的了解，以及電腦作為文本分析工具的出現，適讀性研究者參考閱讀理解的認知研究成果，在適讀性模式中納入一些可能影響閱讀理解歷程的文本屬性；這個階段被稱做認知理論期。當文本適讀性模式牽涉的文本屬性愈來愈多，且複雜時，一些比統計迴歸模式更有效率的文本適讀性計算方式受到了研究者的重視；這個階段可被稱做統計語言模式期。因為一些適讀性模式的計算複雜度超出一般使用者的理解範疇，有研究者提出，這可能會影響他們接受適讀性預測模式的意願。本研究的目的如下：(1) 根據建構適讀性模式時，輸入資料的來源是否與適讀性文獻相關（本文稱之輸入透明度），以及適讀性模式之參數透明度，區分出四種建構適讀性模式的取徑。其中，低輸入透明度/高參數透明度的模式是一種新的文本適讀性的預測取徑。本研究將實際建構這四種模式，並比較它們對文本適讀性預測的有效性。(2) 研究顯示，句法複雜度會影響讀者的文章閱讀，但因為這個屬性不易被量化，所以之前的文本適讀性模式均未納入能直接反應句法複雜度的屬性。本研究根據小學課本中的文本，建構各種語式的出現頻率常模，然後據以找出兩個能反應句法複雜度的屬性，並探討能否將它們納入適讀性模

式中。(3) 本研究針對一群現任小學教師，進行問卷調查及訪談，收集他們對幾種適讀性模式之建構取向的看法。(4) 本研究開發了一套診斷式適讀性系統；該系統除了預測文本的適讀性外，還能為使用者提示哪些文本屬性影響文本的適讀性。本研究亦透過問卷調查與訪談結果瞭解國小教師對此系統的看法，調查結果顯示此系統可幫助教師理解及改變文章困難之處。

關鍵字：適讀性、文本屬性、輸入透明度、參數透明度、文本分析、
診斷式系統

Readability Diagnosis System: A Comparative Study of Four Models on Elementary School Textbook

Yu-Hsiang Tseng



Abstract

Reading is an essential process through which people learn and communicate. In order to predict how comprehensible texts were for readers, readability studies identified text properties and build predicting models. Three genres of studies were differentiated in literatures basing on their text properties considered and modeling methods. In the genre of traditional methods, researchers used easily available text properties (e.g. percentage of difficult words, sentence length, etc.), and used models such as linear regression to predict readability. In the genre of cognitive science-inspired methods, readability studies started to incorporate theories from cognitive science and include more text properties relating to reading comprehension. Some of these properties could be extracted by computerized automatic text analysis tools. As models involved more and more properties, genre of statistical language modeling methods emerged. Researchers employed more elaborate models to predict readability. However, other researchers argued these elaborate models were not easily understandable by average users, and therefore affected how users would adopt the predictions.

Purposes of current study were as follows: (1) four modeling approaches were differentiated by input transparency, which based on the relationship between model's input data and readability literatures, and parameter transparency. Among them, a new modeling approach (i.e. the one with topic modeling) of low input and high parameter transparency was attempted to predict readability of text. This study implemented four

models and compared their performances on predicting readability. (2) Literatures showed that text properties related to syntactic complexities affected reading comprehension, but these properties are not easily computable and thus few readability models directly incorporated them. This study built a frequency norm of phrasal patterns, based on which text properties were extracted to reflect syntactic complexities of phrases. These properties were then tested if they improved readability models. (3) In a survey study, we interviewed teachers in elementary schools, and collected their opinions on how willingly they were to adopt predictions made by different readability modeling approaches. (4) A readability diagnosis system was developed. The system not only predicted readability but provided extra information on the properties affecting readability. Survey studies on teachers showed the diagnosis system could help them understand text properties and edit text.

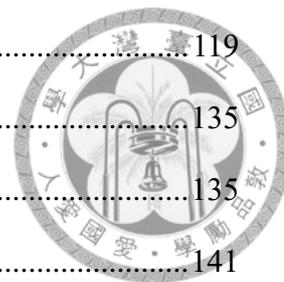
Keywords: readability, text properties, input transparency,
parameter transparency, text analysis, diagnosis system

目次



第一章 前言.....	1
第一節 早期適讀性研究.....	4
第二節 認知心理學取向的適讀性研究.....	12
第三節 文本特徵與自然語言處理.....	29
第四節 診斷式的適讀性系統.....	39
第二章 研究一、文本屬性分析.....	55
第一節 文本屬性的抽取與分析.....	55
第二節 文本屬性與適讀年級迴歸分析.....	68
第三章 研究二、四種取向建構之適讀性模式 的開發與比較.....	77
第一節 研究方法.....	77
第二節 結果與討論.....	84
第四章 研究三、四種不同模式信任度比較.....	95
第一節 研究方法.....	95
第二節 結果與討論.....	97
第五章 研究四、建立適讀性的診斷系統.....	99
第一節 系統架構.....	100
第二節 系統功能.....	101
第三節 文本診斷與編輯.....	106
第六章 研究五、評估診斷系統.....	109
第一節 研究方法.....	109
第二節 結果與討論.....	111
第七章 綜合討論.....	113

第八章 參考資料.....	119
第九章 附錄.....	135
第一節 語式常模.....	135
第二節 文本屬性列表.....	141
第三節 文本屬性與年級常模.....	143
第四節 文本屬性的迴歸分析表.....	147



表圖目次



表 2.1 各模式在各文本類型中之迴歸分析結果	73
表 2.2 完整模式中顯著預測適讀性之文本屬性	75
表 3.1 四種適讀性模式	77
表 3.2 四種模式之交叉驗證結果	91
表 3.3 各適讀性模式正確率	92
表 6.1 診斷式系統問卷之題目列表	110
圖 2.1 句法剖析結果	62
圖 2.2 國語課本語式頻率與長度分布	66
圖 2.3 社會課本語式頻率與長度分布	66
圖 3.1 各模式在不同文本類型下預測之混淆矩陣	92
圖 3.2 各模式預測在不同文本類型下之相關矩陣	93
圖 4.1 小學教師對四種不同模式願採用程度之排序	98
圖 5.1 診斷式系統架構	101
圖 5.2 診斷式適讀性系統主要介面	103
圖 5.3 文本屬性畫面	104
圖 5.4 文本難字診斷	104
圖 5.5 文本難詞診斷	105
圖 5.6 文本難句診斷	105
圖 5.7 與輸入文本相似的文章	106
圖 5.8 文本診斷畫面	107



第一章 前言



讀者藉由閱讀從文本中獲取其所承載訊息。透過分析文本，研究者能客觀地理解讀者從文本中可能獲得的訊息，甚至如何獲取訊息。早在 19 世紀末，一位在大學教授英語文學的教師 Sherman 就認為對文本客觀、科學地分析，有助於學生欣賞、理解文學作品的精神，甚至更能讓學生具體的欣賞、批判文學作品。在他的著作中，他區分出文學作品中的各個元素，例如每個小節（stanza）的用字數、詞彙在朗誦時的力度等等，甚至以圖表的方式呈現這些資料，讓學生理解詩歌所傳達的精神（Sherman, 1893）。

從 Thorndike（1921）在 *Teachers' Word Book* 一書開創對文本適讀性（readability；下文將以適讀性稱之）的研究，以及 Lively 與 Pressey（1923）提出第一個適讀性研究至今，已將近百年。雖然這個領域的研究者大都承襲上述 Sherman 對文本進行客觀分析的想法，並用類似 Thorndike 根據文本的屬性推算適讀性，但 Benjamin（2012）還是根據適讀性模式的建構方法，以及模式所用之文本屬性¹的考量，將適讀性研究分成三個時期：（1）傳統公式期、（2）認知理論期，以及（3）統計語言模式期。也就是說，適讀性研究在「計算預測」（傳統公式時期和統計語言模型時期的主要考量）與「認知解釋」（認知理論時期的主要考量）兩個因素上，漸漸的改進適讀性公式的建構方式。

在傳統公式期沒有電腦輔助文本分析，所以這時期的研究者大都找一些既能預測文章難度（適讀性），又容易取得的文本屬性。往往，這些屬性都是和文

¹ 本研究以文本屬性（properties）代表在文本中，以人工或電腦自動抽取出的文本相關數值（如音節數、字頻、句長、名詞比例、句法樹深度等數值）。這些文本屬性可用在傳統統計迴歸模式中作為迴歸變項（variables），或在其他機器學習模式（如支持向量機）中作為特徵（features）。在本研究中，部份模式是以文本屬性作為輸入，其他模式則直接以文本詞彙作為輸入。故本研究將所有模式的輸入都泛稱為「輸入資料」（input data）。

章內容以及語法複雜度無關的文本表層屬性（如，文本中含有的「難詞」詞數，句子的長度等）。例如 Thondike 就建立一個教師詞彙表，提供一萬個英語詞彙在一些讀物中的使用頻率。教師能夠以此作為工具，評斷一篇文章所包含詞彙是否適合學生閱讀。在接下來的數十年間，眾多研究以詞彙難度做基礎，發展出許多不同的適讀性模式，來預測文本的難度或適讀年級；其中以線性迴歸模型建立模式的方式最被學界採用。

其後，受到認知心理學思潮的影響，「認知解釋」的想法逐漸萌芽，研究者開始探討文本的適讀性與閱讀理解歷程間的關係，並企圖從文本中挖掘更多影響閱讀理解歷程的文本屬性，並根據這些屬性建構適讀性模式。此時期就是認知理論期。但是，因為這些屬性（如，一篇文章中的文字和內容之間的相關性，以及句子的句法複雜性等）不易量化，所以鮮有研究能夠將閱讀的認知研究結果真正落實於具體的適讀性模式中。近年，也就是 Benjamin 所說的統計語言模型期，結構與非結構化資料的快速累積，機器學習演算法迅速發展，研究者開始應用許多分類器來分類文本。例如，從一些已經被標定年級或難度的文本中，應用監督式學習讓電腦得以預測文本的適讀性。這個取向的研究讓適讀性模式的建構者找到更多的屬性、更有效的演算法，建構出的預測模式能更準確預測資料集裡的文本的適讀性。

雖然學者將適讀性研究的發展分為三個時期，但這並不意味著，一個時期的出現就取代了上一個時期。事實上，即便在認知時期的應用研究中，傳統的適讀性公式在近年仍然是重要的研究工具。例如，在一個探討因為車禍而民眾傷亡的研究中，就發現造成幼童傷亡的一個重要原因是父母沒有正確的在車上安裝幼兒椅（Wagner & Girasek, 2003），而造成他們沒有正確安裝的原因可能是「安裝手冊」撰寫的太難。研究者以適讀性公式計算安裝手冊的適讀年級，發現其適讀性多在 10 年級以上。所以研究者建議，廠商應該重新撰寫安裝手冊，降低其適讀年級。



用適讀性公式改變文本的難度，以幫助讀者對文本的閱讀理解，一直是適讀性研究很重要的應用目的之一。Klare (1976) 就回顧了 36 個以適讀性公式改變文本難度的研究。但是，在這些研究中，僅有 19 個研究顯示適讀性公式在改變難度上有正面幫助，另外 17 個研究卻未發現正面效果。Klare 認為，那些未能發現正面效果的研究，一部分是因為情境因素（如讀者的動機、作業設計等），另一部份則是和研究者如何改變文章難度有關。他認為改變文章難度時，不僅要考慮適讀性公式有用到的屬性（例如詞彙難度和句長），也應該考慮功能詞彙的比例、詞的具體性、語式（如主動被動）、句子的複雜度等屬性。只是，或許侷限於當時客觀條件的限制，當時的研究無法用方便有效的方式從文本中提取出這些文本屬性，所以這類的考量難以實際落實到適讀性模式。近年來，統計語言模式期雖提供更多更準確的適讀性預測方法，但其模式的預測方式不能直接讓使用者瞭解文章難度，或者幫助使用者編輯文章難度。

本研究的目的即希望（1）比較幾種適讀性模式的建構取向在文本適讀性的預測表現，也希望（2）透過實際開發一套診斷式系統，該系統除了預測文本適讀性外，亦能指出是哪些文本屬性，影響這篇文本的適讀性。本研究期待此診斷式系統的設計能讓使用者（如，教師）因為適讀性模式可幫助他們理解文章難度，甚至編輯文章難度，進而提升他們接受適讀性模式之計算結果的意願。

文本適讀性研究在 20 世紀初就開始蓬勃發展。為呈現適讀性研究豐富的研究脈絡，本論文在此章第一節首先回顧 1920-1960 年時期的早期適讀性研究。其次，第二節將討論受認知心理學的閱讀研究影響的文本一致性指標。第三節將探討自然語言處理如何幫助適讀性研究者自動地抽取文本屬性。當自然語言處理自動地抽取出文本屬性，研究者面臨的挑戰是如何整合這些屬性並做出預測，所以第四節將討論各種不同的預測模式之間的取向，以及這些取向如何影響到輸入及參數透明度。

在接下來的論文中，研究者根據（一）適讀性模式的輸入資料與文本難度



(適讀性)之間的關係，區分出「輸入透明度」。例如，在適讀性文獻中指出「文本包含低頻字的比率」該屬性是與文本難度相關的，所以使用該類屬性為輸入資料的模式，具備較高的輸入透明度；相反地，有些模式以詞彙序列作為模式輸入，這在傳統公式期與認知理論期的文獻中未提及，故其輸入透明度較低。(二)根據適讀性預測模式中的參數與文本適讀的關係，區分出「參數透明度」。例如，以迴歸模式預測適讀性時，其係數的意義是各文本屬性之線性組合的權重，此參數與適讀性的關係是較清楚的，故其參數透明度較高。此研究按照高(低)輸入及高(低)參數透明度，區分出四類適讀性模式的建構取向。本研究將實際的建構出這四類的適讀性模式，然後比較它們預測文本適讀性的準確度。

為了探討一般人對適讀性模式的看法，本研究以小學教師為對象，採用問卷和訪談的方式收集小學教師對上述四種適讀性模式的信任度以及他們在教學中的實際使用文本適讀性的可能性。最後，本研究將呈現一個能整合研究結果的診斷式適讀性系統。

第一節 早期適讀性研究

壹、適讀性公式的發展

在一篇回顧整理早期適讀性研究的文章中，Klare (1976) 指出，當時的適讀性模式的建構流程如下；(一)找出幾個語言(文本)屬性。(二)選取多篇內容不同的文本，然後計算這些屬性與這些文本適讀性指標間的相關。文本適讀性指標可以用學生閱讀文本後，理解測驗或記憶測驗的得分，或其它的客觀指標。(三)選取相關最高的屬性，並根據這些屬性建構文本適讀性的模式。常用的建構方法是運用統計迴歸法建構一條適讀性公式。(四)選取另一批文本，

驗證適讀性模式對文本適讀性指標的預測效度。

也就是說，傳統公式期的研究者認為，一篇文章的難度能夠透過幾個與文本有關的屬性計算出來。從現在的眼光看，這些屬性都是一些文章的「表層」特徵，如詞彙難度或句長等。例如，Fry (1977) 的做法是，隨機在一篇文本中選取一段 100 字的章節，然後根據其中所含音節數及句數推算該篇文章的適讀年級。他設計了一個由「音節數」和「句數」兩個屬性為向度的「適讀性圖」，使用者只要按圖索驥，便可以得知一篇文章的適讀年級了（見 DuBay, 2014）。也有研究者則利用一些文本屬性為預測變項，經過線性迴歸分析，即可以一條公式的形式，計算出一篇文章的適讀性。在當時，這種作法也引來許多批評。因為適讀性公式只是透過文本屬性來「預測」閱讀時的難易程度，並無法「測量」閱讀時讀者和文本之間的關係。但是，透過公式計算文本適讀性的方法，比找教師或作者主觀認定文本的難度，適讀性公式顯然具備較高的效度，所以這些研究的結果在當時的教育、軍事、商業等領域受到廣泛的重視與應用（Klare, 1963）。甚至，在多年後的今天，一些適讀性公式仍普遍被用來評估兒童讀物的適讀性，甚至一些商用的文書處理軟體裡（例如，微軟文書處理軟體 Word）會內建適讀性的公式（如，Word 即內建 Flesch 適讀性公式）。

根據 Klare (1976)，文章難度的客觀校標是建構適讀性模式的重要依據。因為在模式開發期，研究者需要找到一組難易不等的文章，然後藉著分析這些文章的文本屬性，找出哪些屬性能夠預測文章的難度。當適讀性模式建構出來以後，研究者又需要找出另一組難易不等的文章，來檢驗適讀性模式的預測效度。早期的適讀性研究的文章難度校標，多使用讀者在閱讀理解測驗、記憶測驗，或是填空測驗的表現（Klare, 1974）。例如，McCall-Crabbs 標準閱讀測驗（McCall-Crabbs Standard Test Lessons in Reading）（McCall & Crabbs, 1961）的作法是，收集多篇文章，根據文章的行文及內容設計一些選擇題，然後請程度不等的讀者（如不同年級的學生）閱讀文章並作答。根據讀者閱讀一篇文章

後，對測驗題目的答題表現，研究者可以推斷這篇文章適合哪個年級的學生閱讀。例如，閱讀一篇文章後，三年級學生平均能答對一半以上的題目，這篇文章的適讀性是三年級。也就是說，這是利用讀者在測驗題目上的表現，評估一篇文章的難易程度；這在當時是很常用的文章難度效標。



另外，將文章設計為填充測驗 (cloze test) 也是一種測驗文章難度的方法。在填充測驗中，一篇文章有某些字或詞被設定為空格 (如，每 5 個字/詞，就有一個被掩蓋)，讀者的作業是猜測，並填入空格中的字。為了控制被掩蓋之字在句子中的位置對讀者答題的影響，研究者可以設計不同的填充測驗的版本；例如，在一個「逢五刪一」的測驗中，研究者至可以設計 5 個不同的版本；版本 A 將第 1、6、11... 詞為空格；版本 B 為第 2、7、12 詞為空格，以此類推。Bormuth (1969) 做了一個大規模的適讀性研究。他用讀者對文章填充測驗的表現作為文章適讀性的客觀效標。在研究中，Bormuth 選了 330 篇長度約 100 詞的文章，文章中每 5 個詞，留空一詞；每篇文章依據空格位置，設計出 5 個版本。研究參與者是 2,600 位 4 到 12 年級的學童；Bormuth 用參與者在填充測驗上的表現，估計這 300 篇文章的難度。在諸多評量文章難度指標的方法中，填充測驗似乎是較為適讀性研究之研究者青睞的方法。例如，Miller (1974) 發現，填充測驗具備較能完整涵蓋文章難度的優點，而且以填充測驗的分數作為適讀性公式的效標，也能得到較好的預測效果。又如在第一個使用填充測驗作為驗證適讀性公式校標的研究中，Coleman (1965) 發現，適讀性公式的預測與填充測驗的結果之間的相關最高可達 .91；比以 McCall-Crabbs 分數作為效標，得到的相關高 (Szalay, 1965)。

根據 Klare (1976)，尋找與文本難度相關的屬性是建構適讀性模式的另一個重要步驟。在適讀性研究發展的初期，研究者認為文章中的「詞彙負荷」 (vocabulary burden) 是造成文章難度的重要因素。例如 Lively 與 Pressley (1923) 認為國中的科學教科書充滿艱澀的技術性詞彙，使得學生難以學習。

於是他們的適讀性計算每 1000 個詞中不同的詞彙數量，以及在這些詞中，不在 Thondike 詞表裡的詞數。他們發現用這個簡單的屬性，可以和 700 本書的難度達到 0.80 的相關 (Dubay, 2004)。適讀性和「難詞」有關的想法，影響到後續的適讀性公式研發。例如：Lorge (1939) 提出的適讀性公式可用來篩選 3-12 年級的適讀文本 (Klare, 1974)，該公式中用了三個屬性：

$$\begin{aligned} X1(\text{年級程度}) &= 0.06 \times \text{平均句長 (詞數)} + \\ & 0.10 \times \text{每 100 詞中的平均介系詞片語數} + \\ & 0.10 \times \text{難詞數 (不在 Dale 詞表中的不同詞數)} + \\ & 1.99 \quad (\text{常數項}) \end{aligned}$$

以 McCall-Crabbs 測驗做驗證此適讀性公式時，得到頗高的相關。

Lorge 公式裡的難詞定義來自於 Dale 的 769 詞詞表。該詞表是 Dale (1931) 根據 Thondike 詞表改編而來。Dale 和 Chall (1948) 再次修訂該詞表，列出 3000 個簡單詞，並提出 Dale-Chall 公式；此公式與 McCall-Crabbs 測驗分數的相關達 0.70。Dale-Chall 公式如下：

$$\begin{aligned} Xc50(\text{年級程度}) &= 0.1579 \times \text{難詞數 (不在 Dale 詞表中的不同詞數)} + \\ & 0.0496 \times \text{平均句長 (詞數)} + \\ & 3.6365 \quad (\text{常數項}) \end{aligned}$$

然而，有些研究者認為，應有其他更基本的詞彙屬性可描述詞彙難度。例如 Flesch (1943) 認為詞彙的抽象性和詞彙的難度有很重要的關係，而他的公式中就用詞綴 (affix) 個數當作衡量詞彙抽象性的指標，他的公式如下：

$$\begin{aligned} \text{年級程度} &= 0.1338 \times \text{平均句長 (詞數)} \\ &+ 0.0645 \times \text{綴詞個數} \\ &- 0.0659 \times \text{個人指涉個數} \\ &- 0.7502 \quad (\text{常數項}) \end{aligned}$$



公式中的「個人指涉詞」指的是（在英文中）人名、代名詞（如 he、she）、人感興趣的詞（如，man、woman、child、boy、girl 等）。該公式也同樣使用 McCall-Crabbs 測驗做驗證，並得到高相關。

在當時，由於沒有其他工具的輔助，一篇文章的適讀性完全需要人工計算。在上述的公式中，計算詞綴個數需要一個個詞檢查，讓該公式的實用性大幅降低，同時，也因為計算個人指涉的定義較為模糊，於是 Flesch（1948）提出一個新的「閱讀簡易度公式」（reading ease）。該公式同時也是微軟文書處理軟體（Word）到目前為止仍提供的適讀性公式：

$$\begin{aligned} \text{Reading Ease} &= 206.835 - 0.846 \times \text{每 100 詞中的音節數} \\ &- 1.015 \times \text{平均句長 (詞數)} \end{aligned}$$

如同 Flesch 在 1943 年公式中的詞綴數，Flesch 認為音節數也是可代表詞抽象度的指標。Farr、Jenkins 和 Paterson（1951）考慮到計算音節數需要公式使用者知道音節規則，他們發現若只使用單音節詞的個數來替代音節數，也可以達到同樣的預測效果。Farr 等的公式與 Flesch 的公式相關高達 0.93-0.95。他們提出的新閱讀簡易度公式是：

$$\begin{aligned} \text{New Reading Ease Index} &= 1.599 \times \text{每 100 詞中的單音節詞數} \\ &- 1.015 \times \text{平均句長(詞數)} - 31.517 \end{aligned}$$



Gunning 的 Fog Index 也同樣用音節數來衡量難詞 (Gunning, 1952)。他提出用三個以上音節的詞數用來當作「難詞」的指標。McLaughlin (1969) 則認為計算平均句長需要較多時間，於是他提出更容易計算且準確率更高的 SMOG 適讀性，SMOG 只用到多音節詞數 (三個以上的音節)，和句數。這兩個適讀性公式是：

Gunning's Fox grade level = $0.4 \times (\text{平均句長} + \text{三或多音節詞數比例})$

$$\text{SMOG} = 1.0430 \cdot \sqrt{\text{多音節詞彙數} \times \frac{30}{\text{句數}} + 3.1291}$$

由於適讀性公式的發展動機與應用場域的關係，讓適讀性公式容易使用一直是適讀性研究的趨勢 (Klare, 1974)。例如，Danielson 與 Bryan (1963) 首次嘗試讓電腦計算適讀性公式所需要的屬性。Danielson 與 Bryan 公式含有詞長和句長等兩個屬性，但為了讓電腦容易計算，他們改用「字母」(character) 作為計算單位：

$$\text{DB Score} = 131.059 - 10.364 \times \text{空格間含有的平均字母數} \\ - 0.194 \times \text{一個句子含有的平均字母數}$$

為了讓適讀性公式更能廣泛應用，FORCAST 閱讀等級 (RGL) 公式僅包含一個屬性 (Caylor, Sticht, Fox, & Ford, 1973)；這個屬性是，每 150 個詞中，含有幾個單音節詞。該公式如下：

$$\text{FORCAST RGL} = 20 - \frac{\text{單音節詞數}}{10}$$



因為 FORCAST 的方便性，美國軍方曾經使用它來評估文本的適讀性。

另一種增加適讀性公式易用程度的方法是 Fry (1977) 的適讀性圖。該適讀性圖的 x 軸代表每 100 個詞的音節數，y 軸則是每句的詞數。讀者只要依照 x 軸和 y 軸的數值比對圖形，即可得到適讀性結果。該適讀性圖在 1980 年代間受到非常廣泛的應用 (Klare, 1988)，其效度和其它適讀性公式的結果也非常相似 (Klare, 1963)。

過去的中文適讀性研究中也曾試圖以詞彙、字數等屬性預測中文文章適讀性。例如陳世敏 (1972) 就用句子的平均字數和難字的多寡來預測中文文章的適讀年級，Yang (1970) 則藉常用詞彙數目、字彙的平均筆畫數以及含有主詞及述詞的句數。荊溪昱 (1995) 以句子長度、課文長度、常用字比例為預測屬性，預測國立編譯館版本的國小及國中教科書的年級。胡志偉、方文熙及李美綾 (1994) 參考 Klare (1976) 的適讀性模式建構流程，先以文章難度排序作業、詞填充測驗，以及字填充作業評估 21 篇短文 (文章長度約 200 到 400 字) 的難度。其次以這三種不同測驗結果為效標，用多元迴歸的方法從 23 個文本屬性中，篩選少數幾個屬性，建構適讀性公式。研究發現，「文本包含的長子句數目」這個屬性就可以解釋排序作業 55.14% 的變異量；「長子句數」和「高頻詞數」可解釋詞填充作業 48.89% 的變異量。

貳、傳統適讀性公式的批評

前文臚列的數條適讀性公式，只是適讀性文獻的一小部份。DuBay (2004) 指出，在 1980 年代時，研究者已經提出超過 200 條的適讀性公式。雖然研究結果豐碩，但這些適讀性公式卻接受到來自兩個層面的批評：(1) 適讀性公式僅能「反應」文章難度，但無法「創造」出適合閱讀的文章；(2) 這些

公式使用文本的表層屬性預測適讀性，而文本表層屬性無法反應閱讀時的深層認知機制。(Benjamin, 2012; Bormuth, 1966; Davison & Kantor, 1982)。

關於第一項批評，研究者確實發現，適讀性公式難以讓作者靠著適讀性公式「寫出」一篇好文章。甚至當作者刻意「順著」適讀性公式的評估結果寫作，反而會寫出不易閱讀的文章 (Schrivver, 2000)。這個批評的確影響到適讀性的應用層面，但 Klare (1976) 認為適讀性公式的目的是預測適讀性，而不是因果性的提出修正建議，所以他區分兩種適讀性的研究取向；其中一種是產生可讀的文章，另一種是預測可讀的文章。前者適讀性公式中使用的文本屬性必須是有因果關係的，且必須透過實驗法驗證其有效性；但後者才真正是適讀性公式的研究目的。適讀性公式應該是用作為一種大規模的「過濾工具」

(screening)，它使用到許多指標性的文本屬性來預測適讀性，並透過相關法來確認其預測的有效性。

第二項批評則來自於適讀性公式使用之屬性的本質。由於適讀性公式是預測性的文本篩檢工具，所以其包含的各種屬性，無論其本質為何，只要能夠幫助適讀性公式提升它對文本客觀效標的預測效度，且容易計算，就是適讀性公式會使用的屬性。有些研究者則質疑，適讀性公式常用的音節數、句子長度等屬性，難以關連到真正適讀性所聲稱要測量的閱讀難度。亦即，適讀性所使用的文本屬性缺乏表面效度 (Schrivver, 2000)。

另外，也有研究者質疑，適讀性研究常用的文本難度效標不夠精確。Stevens (1980) 指出，McCall-Crabbs 的標準閱讀測驗並不夠嚴謹。Kintsch 與 Vipond (1979) 質疑填空測驗的分數反映的可能是文句中的訊息重複性 (redundancy)，而不是閱讀難度。

這些對於適讀性公式的批評，反映研究者對適讀性模式的省思。尤其，對適讀性公式所使用之屬性的本質的討論，更反映研究者在建構適讀性模式時，不單只考量「預測」，也加入了「認知」的考量。傳統的適讀性公式就像是以

「頭髮長度測量性別」。它們的確取得某種程度上的正確率，但只考慮到文章中的表層訊息，難以真正描繪讀者閱讀背後的心理歷程（Miller & Kintsch, 1980）。隨著當時認知心理學的發展，適讀性研究者開始注意「認知」因素，試圖讓適讀性研究更能呼應認知心理學中的閱讀理論。



第二節 認知心理學取向的適讀性研究

認知理論期的適讀性研究反映了當時認知心理學的蓬勃發展，促使適讀性研究在「計算預測」，在適讀性模式中反應人類閱讀的認知歷程。認知心理學在1950-1960年代萌芽，並在隨後的數十年間蓬勃發展。它與過去行為主義心理學最大的不同，在於研究者不僅研究行為，更研究行為背後的心理機制。在閱讀研究中，除了過去操弄文本的表層文字屬性（如字體大小、排版等）外，認知心理學者進一步探討人們閱讀時，所牽涉的心理機制。例如，一篇文本含有的語言層次，以及處理各語言層次的訊息時牽涉的心理歷程。

閱讀是讀者藉由視覺管道從文本中獲取訊息，並理解文本意義的歷程（Rayner, Pollatsek, Ashby, & Clifton, 2012）。這個定義涵蓋了許多與閱讀歷程有關的研究。例如，研究字詞辨識歷程的研究者探討人們如何從接受視覺刺激（詞或是字）到從刺激中取得字形、字音和字義的歷程。他們在實驗中將字（或詞）單獨呈現或呈現在語境脈絡中，然後操弄其各種與刺激字或語境脈絡相關的屬性，最後通過參與者對刺激字的反應推論字彙辨識的心理機制。對句子結構或句法與語義互動感興趣的研究者則會操弄句子的句法特徵及語義特徵，並以句子（sentence）作為刺激材料，然後從參與者對句子的反應推測人們閱讀句子時的心理機制。這些研究提供適讀性研究豐富的理論基礎，亦即讀者在閱讀時如何理解一篇文章；以及哪些文本屬性，可能會影響讀者的閱讀理解歷程。

壹、文本的心理表徵



有關閱讀理解歷程的理論頗多，本文將引用其中一個已經將部分想法實際落實於適讀性模式的理論（Graesser, Singer, & Trabasso, 1994; Kintsch, 1988）。

因為這個理論的一個目的是建立一個能夠處理語言的人工智慧系統，所以它的部分看法不一定被所有的語言研究者接受（如 McKoon & Radcliff, 1992）。根據 Kintsch 的理論，閱讀時，讀者會在文本中主動的搜尋訊息，以便他們建立文本內容的心理表徵。該表徵需具備「一致性」（coherence），亦即每個元素（字詞、子句、句子）間都必須彼此連結，成為一個協調的組織，且不同句子所形成的組織，也需要相互整合成更高層次的組織，最後，這些組織將被抽象為文章的主旨、綱要等。這樣的心理表徵就是讀者對文章的理解。當讀者能夠成功理解一篇文本時，讀者就能進行與文本內容相關的推論，也能根據文本提出問題，例如指出文本內容的邏輯錯誤、矛盾或缺漏等。同時，讀者也能正確回答關於文本的問題，也能精準地摘要出文本的訊息（Graesser et al., 1994）。

讀者對文本建立的心理表徵包含 5 個層次：（1）淺層編碼（surface code）；（2）文本表徵（textbase）；（3）心智或情境模式（mental/situational model）；（4）文本類別與修辭結構（genre and rhetorical structure）；（5）語用及溝通目的（pragmatic communication）（Graesser, McNamara, & Kulikowich, 2011; Graesser, Millis, & Zwaan, 1997）。讀者需有足夠的認知資源和對語言及世界的背景知識，才能成功建立上述的心理表徵。以淺層編碼的訊息為例，該層次的訊息包含表層語言編碼的辨識及剖析，包括文字的視覺訊息，聲音訊息，音節、詞素、詞綴、時態（tense）和型態（aspect）。同時也包含詞彙的詞類和語法結構（如名詞、動詞片語）等。在英文的研究也指出，初級的文字編碼階段的確會影響之後的閱讀歷程（Rayner, 1998）。讀者的心理辭典中是否包含足夠的詞彙數量，也會影響讀者的閱讀理解程度（Perfetti, 2007）。例如，初次接觸中文

的讀者，很可能無法理解中文的書寫系統、發音、詞彙、文法規則，則他很可能無法建立初級的表層編碼，可能亦難以繼續建立更高層次的文本訊息。

第二層次是文本表徵。在文獻中，文本內容常以命題（propositions）的形式表達（Graesser et al., 1997; Kintsch & Dijk, 1978）。「命題」在概念上是指涉到外在（真實或想像）世界的狀態、事件或動作，且該指涉具有真假值（truth value）。在文本表徵裡的命題具備一個述詞（predicate），以及一或多個論元（arguments）。述詞在命題中常是句子的主要動詞、形容詞或連接詞，而論元則可能是名詞或其他的命題。例如以下取自 Graesser 等人（1997）的例子：

A mushy, brown peach is lifted from the garbage and placed on the table to pinken.

PROP 1: lift (AGENT = X, OBJECT = peach, SOURCE = from garbage)

PROP 2: brown (OBJECT = peach)

PROP 3: mushy (OBJECT = peach)

PROP 4: place (AGENT = X, OBJECT = peach, LOCATION = on table)

PROP 5: pinken (OBJECT = peach)

PROP 6: [in order] to (PROP 4, PROP 5)

PROP 7: and (PROP 1, PROP 4)

從上述例子可發現文本表徵和表層編碼不必然是相同的。例如，peach 和 table 在命題中都是以同樣的形式表達，並未區分在句子中一個是不定指（用不定冠詞 a），一個是定指（用定冠詞 the）。同時，命題形式僅表徵句子的意義訊息，故已去除原本文句中的被動語態。

文本表徵中的命題形式很可能是斷裂的，讀者需透過各種推論歷程建立文本的「微觀結構」（microstructure），該結構將文本中的命題連成一個完整的網絡。讀者在短期記憶中透過各種策略來推論命題與命題之間的關係，例如命題的論元重複（例如上述的 PROP 2 和 PROP 3 都和 PEACH 有關）、代名詞指



涉、概念重複，或連接詞等，都能幫助讀者建立文本的微觀結構。讀者閱讀完一個段落後，進一步從文本表徵中抽取出巨觀結構（macrostructure）。巨觀結構是讀者透過直觀推理（heuristic），例如藉由長期記憶裡儲存的世界知識來推論，或應用各種文本線索（因果、時間順序、並列等等）來產生命題間的連結，最後經過刪除、概化等歷程，建立較高層次的篇章主題（theme）或綱要（gist）（Kintsch & Dijk, 1978; van Dijk & Kintsch, 1983）。Graesser 等人（1994）曾列舉 13 種在閱讀敘述文體時，讀者用來建立微觀、或巨觀結構的推論類型，除了和語言符號相關的線索外，也包含因果關係、目的、主題、工具目的、敘事狀態等。過去實驗也發現，讀者在閱讀的當下，便建立其所讀文句的意義。例如 Long、Golding 及 Graesser（1992）在實驗中讓讀者每讀完一句話，就出現一個測試詞彙，參與者要猜測該詞彙是該句話的上位或下位（superordinate/subordinate）目的。讀者看到測試詞彙後需盡快唸出它。實驗者測量參與者唸出詞彙的念名時間。實驗結果發現，讀者對上位目的詞彙的念名時間顯著低於下位詞彙和控制組，顯示讀者在閱讀句子時，便理解句子的意義，並根據句子意義進行關於上位目的推論。

微觀結構和巨觀結構都涉及文本內容以及讀者的長期記憶訊息，這反應閱讀時的心理表徵，並非只有單純的文本表徵。這些在閱讀過程中所牽涉的各項訊息，都整合在讀者對文本的情境模式（situation model）裡。情境模式是一個來自文本的新訊息，與讀者舊有的長期記憶間訊息整合的過程。讀者在閱讀時在情境模式中建立命題與命題之間共同指涉或推論關係，讓文本的命題結構與讀者自己的背景知識形成一致性的知識網絡。此心理表徵在過去文獻中曾被稱為可能世界（possible worlds）、篇章模式（discourse models）（van Dijk, 1976）、參照網絡（reference nets）（Habel, 1982）或心智模式（Johnson-Laird, 1980）。過去實驗也指出，讀者對於文本的記憶是讀者建立出的情境模式，而不是淺層文字本身（Garnham, 1987）。讀者從文本中學習新知時，亦是透過舊有

知識的情境模式，並藉由文本的新資訊逐步更新原有訊息。

文章的淺層編碼、文本表徵和情境模式在文獻中有許多實徵研究的支持。例如，在過去研究中，研究者要求參與者閱讀一篇文章，之後需對四種不同的測試句做再認作業（recognition），再認的材料是：(1) 文章中的原始句子；(2) 由原句改寫句子；(3) 由原文中合理推論出的句子；(4) 一個錯誤的句子。研究者以相減法探討不同層次表徵的效果，亦即(1)減去(2)可以得到表層編碼的訊息；(2)減去(3)透露的是文本表徵訊息；(3)減去(4)則代表情境模式的表徵。實驗結果發現，淺層編碼訊息隨著時間間隔快速衰退，但愈高層的情境模式則衰退得最慢，文本表徵則居中（Kintsch, Welsch, Schmalhofer, & Zimny, 1990）。更有趣的是，研究也發現當讀者認為閱讀的文章是文學作品時，表層編碼的表徵較強（衰退得較慢），而當讀者認為文章是新聞報導時，情境模式的表徵則較強（Zwaan & Radvansky, 1998）。這同時呼應了文章類別在理解文本時的效果。

研究發現，當文本中出現為違背語用原則的訊息時，會使讀者閱讀理解的時間變長，而且容易出現理解上的錯誤。例如，Graesser 等人（1997）指出，作者會透過語用訊息和讀者溝通。在寫作時，作者會用與讀者共享的知識寫文章；當作者提出新的訊息時，會善用篇章線索。例如，以英文寫作的作者會將舊的訊息放在主語位置的名詞片語中，而新的訊息通常放在動詞片語中。作者也善用篇章結構以提示重點訊息，比如將重要的訊息放在每段的第一行。另外，Graesser 等人（1997）也延伸 Grice（1975）在溝通中的合作原則（Grice's axiom），認為作者應在情境模式之下提出真實的陳述，同時所有句子都應該和前文相關，文章彼此之間不應互相矛盾抵觸。過去研究發現，一般讀者很難意識到作者的語用線索，尤其是閱讀含有不熟悉議題的說明文持，即便文章中的語用線索有衝突，讀者還是會認為自己已經理解文章了，而忽略文章中的矛盾訊息（Glenberg & Epstein, 1987）。





上述五個層次的文本表徵，在實際閱讀文章時，是互相依存的。雖然讀者若在建立底層表徵時受到阻礙，可能會影響高層次表徵的處理。例如，當一位無法辨識中文字的讀者閱讀一篇中文的操作手冊時，他可能在第一層的文字編碼層次就已受阻，亦無法建立文本表徵甚至情境模式等。然而，高層次的表徵仍有可能透過其他線索的協助，完成該層次的訊息處理。舉例而言，兩位同系的大學部同學共同合作一篇課堂報告。一位同學閱讀過初稿後，雖然他能夠理解整篇報告的內容（因為他們共享相同學系的背景知識），但仍然覺得報告的行文不清楚，例如文中缺少轉折語、連接詞等標示文章結構的用語。在這個例子中，即便文本表徵中沒有提供足夠的線索給讀者，但讀者心中的背景知識（長期記憶）還是能建立情境模式。甚至，讀者可以用情境模式的訊息幫助填補文本表徵缺漏的訊息。

然而，也有些文本層次間的訊息是不能互相補足的。例如，在 Graesser 等人（2011）曾舉的例子中，兩人為了組裝一套新的視聽器材，而試圖閱讀一本說明書。他們有正常的閱讀能力，所以在淺層文字編碼和文本表徵的層次都能正常處理。他們甚至也都完全理解說明書的溝通目的（教使用者如何組裝）和文類修辭結構（說明書是說明文類），但他們就是無法弄清楚到底哪一個線材要接哪個設備（情境模式的訊息）。在這個例子中，這兩個人的淺層和文本表徵訊息無法幫助他們建立情境模式，上層的文類和語用層級的訊息也無助於情境模式的建立。這個例子反應出有些問題只能透過額外的背景知識，才能幫忙補足（rectified or repaired）情境模式的缺陷（misalignment or misfire）。

貳、讀者如何建立一致性的心理表徵

當讀者「理解」一篇文本時，他能建立出一致性的心理表徵。讀者藉由該表徵能處理與文本相關的問題，例如，回憶文本內容、對文本進行摘要、重述



(paraphrase) 文本內容，以及回答與文本內容有關的問題等等。「建構整合理論」描述了讀者如何建立出一致的文本表徵。該理論於 1980 年代提出，但在 2016 年的「促進國際閱讀素養研究」(PIRLS) 的閱讀架構仍深受「建構整合理論」的影響 (Mullis & Martin, 2015)。根據建構整合理論，讀者不但能從下而上，彈性地從文本建構出各種訊息；同時也能用過去的知識和記憶，從上而下地將建構出的訊息，整合進過去的經驗架構中 (Kintsch, 1988)。該理論的建構過程分為四個階段：(1) 從語言內容建立命題結構；(2) 從背景知識提取與文本相關的知識；(3) 從文本中推論出其他命題；(4) 對命題間的關係設定權重。上述四階段僅是建立初步的粗略文本表徵，所以裡面可能包含不一致的心理表徵，或甚至與文本無關的訊息。整合階段的目的是要強化與文本脈絡相關的訊息，並抑制與脈絡無關的訊息。在建構整合理論中，整合階段的歷程是一次又一次的激發循環 (cycle) (Miller & Kintsch, 1980)。每一次的循環都以前一次的循環作為基礎，每一個命題節點都具備自己的激發強度，並用命題與命題間的連結權重傳播激發能量。如果某次循環失敗，亦即有些命題節點無法整合進網絡，則新的命題節點會再從長期記憶中提取出來，並再嘗試一次整合循環。此整合階段將一直持續到所有節點的激發強度穩定為止。

Miller 和 Kintsch (1980) 曾嘗試用 InterLisp 模擬整合階段的運算歷程，並和參與者實際閱讀的閱讀時間、命題回憶數做比較。他們的「微觀結構一致性模式」(microstructure coherence program, MCP) 試圖模擬讀者在理解時，連結命題單元以及擷取長期記憶的歷程。該模式具有 3 個記憶區塊，其中輸入緩衝區 (input buffer) 負責儲存目前正在處理的命題單位，如果該命題單位和工作記憶 (working memory graph) 中的文本表徵共享概念 (或論元)，則該命題就被納入工作記憶的文本表徵中。但如果輸入緩衝區的命題單位無法用上述方法連結進工作記憶的文本表徵，則該模式會嘗試在長期記憶 (LTM) 中搜尋能幫助連結的命題單元。長期記憶裡會儲存所有在文本中曾出現過，但已不在工作

記憶中的命題單元。如果長期記憶搜尋仍然無法接起輸入緩衝區的命題單元和工作記憶的文本表徵，則此模式就判斷該文本產生主題轉移 (topic shift)。工作記憶裡的文本表徵就轉入長期記憶，輸入緩衝區的命題單元則進入工作記憶當作新文本表徵的根節點。



雖然 MCP 模式並沒有用到複雜的演算法，但模式的推論數和長期記憶搜尋數和閱讀時間有顯著正相關，推論數則和回憶表現有顯著負相關。Miller 與 Kintsch (1980) 認為，MCP 模式只模擬了部分的閱讀的歷程。如果要進一步貼近真實的閱讀歷程，則應該要再加入語義剖析器，讓文本能自動轉成命題結構 (在 MCP 裡，命題結構是由研究者手動輸入)；另外也應有建立巨觀結構的演算法。

「建構整合模式」和 MCP 模式都試圖提供一套認知虛擬機器 (cognitive virtual machine) (Pylyshyn, 1984)，試圖用形式化的方法描述人類閱讀的認知歷程。然而，建構整合模式的確提供一套解釋模式，讓研究者瞭解某一文本屬性 (例如文本中的連接詞) 可能會如何影響讀者的理解 (連接詞如何影響一致性)，但是由於這個理論相當複雜，所以它難以用實驗法探討讀者閱讀一篇文章後，會引發哪些心理歷程。一篇「困難」的文章之所以困難可能牽涉到各種因素，例如某個命題無法連結進文本表徵，進而引發長期記憶搜尋，甚至因為工作記憶的限制而無法儲存文本表徵，使得讀者要啟動眼動機制再回視 (saccade) 前一段文字等等。或者，研究者也無從得知讀者的長期記憶中，是否擁有連結命題所需要的背景知識，若一次的長期記憶搜尋失敗，之後又將引發其他的理解機制。閱讀所引發的複雜認知歷程，可能難以從認知虛擬機器的角度，判斷讀者是否真的從文本中建立出一致性的心理表徵。

參、文本一致性指標和 Coh-Metrix



讀者對文本的心理表徵是否有一致性 (coherence) 是內在的抽象構念，難以直接測量。但是「文本」是否提供足夠的一致線索 (cohesive devices)，讓文章本身是一致的 (cohesion)，則是可從文本屬性計算而得的²。Coh-Metrix 即採取此取向，找出一些會影響讀者對文本建立各層心理表徵的文本屬性，並以之作為指標，描述一篇文章是否容易被讀者理解。

Coh-Metrix 希望對文本做多層次的難度和一致性評估，並瞭解在不同層次間的難度和一致性會如何彼此互動及限制，並藉此發展多層次的閱讀理解模式。過去實驗也發現高文本一致性 (cohesion) 的文章會幫助低背景知識的讀者閱讀。有趣的是，這個實驗也同時發現，低文本一致性的文章反而有助於讀者主動搜尋背景知識，進而幫助閱讀理解 (McNamara, Kintsch, Songer, & Kintsch, 1996)。這些研究顯示文本的一致性的確對於閱讀理解有重要關係。

McNamara、Ozuru 及 Floyd (2011) 的研究也得到類似地結果。他們直接操弄文本的一致性程度對四年級學生閱讀的影響。研究中使用 2 篇記敘文、2 篇說明文作為閱讀材料，並以 Coh-Metrix 指標來操弄文本一致性，例如把代名詞換成名詞片語、加入可以增加概念指涉性的詞彙，加上句子連接詞等等，以增加文本的一致性；當然，反向操作則會減低文本一致性。研究以根據文本內容設計的選擇題測量讀者對文本的理解程度，結果發現高一致性的文本能增加學童對記敘文的理解程度。他們也發現和 McNamara 等人 (1996) 類似地結果，亦即當學童對於文章的背景知識高時，閱讀低一致性文本反而有更好的閱

² 心理表徵一致性 (coherence)、文本一致性 (cohesion)、以及一致性屬性 (cohesive devices) 是彼此相關聯的概念，但這三者 in 中文翻譯時不易區分。本研究實際操作並計算一致性屬性，並作為預測文章適讀性的屬性之一；該屬性會影響文本的一致性，並進而影響讀者閱讀時建立之心理表徵一致性。

讀理解表現。

Coh-Metrix 1.0 (Graesser, McNamara, Louwerse, & Cai, 2004) 分析超過 200 個文本指標，這些指標分成 13 類，包括詞彙訊息、詞頻訊息、詞類標記、詞彙密度、邏輯詞彙、連接詞、詞類詞例比、多義詞與上位詞、句法複雜度、適讀性公式、共同指涉一致性、因果一致性和 LSA 語意訊息。由於 Coh-Metrix 指標相當豐富，Graesser 等人 (2011) 為凸顯指標與指標在真實文本間的關係，故選擇 53 個指標分析 TASA 的 37,520 篇文章，並用主因子分析法 (PCA) 得到 8 個主要因子；這些因子共可解釋 67.3% 的變異量。他將這八個主要因子分別對應到五個文本層次，與讀者在閱讀時建立的文本心理表徵相當類似 (Graesser et al., 2011)，分別是詞彙、語法、文本、情境模式與文章類型。

一、詞彙層次指標

詞彙層次指標包含詞頻、詞類、熟悉度、具體性、習得年紀 (age of acquisition, AOA) 等。研究顯示，這些屬性都會影響讀者對語文材料的處理時間與理解程度 (Perfetti, 2007)。對閱讀初學者而言，口語化的詞彙較適合他們閱讀者；但對精熟的讀者，文本中含有較複雜的詞彙可能更能引起他們的閱讀興趣。在過去研究中，詞彙的使用頻率是最穩定地影響閱讀的屬性。研究者可以透過大型語料庫 (例如，Coh-Metrix 使用的 CELEX 語料庫) 計算詞彙頻率，並將此詞彙頻率用來估計讀者對其熟悉程度。

詞彙可區分為實詞 (名詞、動詞、形容詞、副詞等) 和功能詞 (介系詞、冠詞、代名詞等)，Coh-Metrix 計算不同詞類在文本中所佔的比例，因為過去研究發現實詞和虛詞在文章中扮演不同的角色，均可能影響閱讀理解。例如，在眼動研究發現，讀者對實詞的凝視時間高於虛詞 (柯華蕨、陳明蕃、廖家寧，2005；Rayner & Duffy, 1986)。然而，功能詞在文本理解中擔負非常重要的角色，例如代名詞負責提示讀者前後文句之間的關係，有助於提醒讀者連結文本中地不同命題單元。同樣地，連接詞詞類也是建立文本一致性時很重要的線

索，例如因果詞（因為、所以）、時間詞（然後、之後、之間）、邏輯關係（如果、因此）、並列（同時、而且）和轉折（然而、但是）等等，這些詞彙也是讀者建立情境模式的線索。



Coh-Metrix 也藉助 WordNet (Fellbaum, 1998) 描述英文詞彙的語意內容 (semantic content)。WordNet 在 Coh-Metrix 裡提供三項關於詞彙意義的訊息。

(一) 多義詞：一個詞彙在 WordNet 的詞義數 (sense) 愈多，代表一個詞彙愈多義。過去文獻指出，多義詞在閱讀時會造成閱讀速度變慢，對初學者的影響尤其深遠 (Gernsbacher & Faust, 1991)。(二) 具體性：WordNet 不僅列出詞彙的意義，也描述每個詞義之間的上下位關係 (hypernym/hyponym)。例如，座椅 (seat) 在 WordNet 中作為名詞時，有 9 個詞義 (sense)，其中一個詞義是「用來坐的家具」(furniture that is designed for sitting on)，其上位詞義 (hypernym) 是家具 (furniture)³。家具的上位詞 (及這個詞的上位詞，以此類推) 依序是裝潢 (furnishings)、工具 (instrumentality)、人造物 (artifact)、物件 (object)、物體 (entity)。一個詞彙的上位詞數可代表其具體性 (上位詞愈多則該詞彙愈具體)；(三) 因果性：Coh-Metrix 也使用 WordNet 的詞彙訊息來定義因果性動詞 (causal verb, 如 kill 等)，這些動詞是在情境模式中，讀者用來理解事件發生順序和因果關係的線索 (Zwaan & Radvansky, 1998)。

二、句法層次指標

句法 (syntactic) 層次指標包含句子的複雜度以及句子與句子間的句法相似度。有些句子具有很簡單的句法形式，例如單純的主動賓 (SVO) 形式 (he went to school；我吃飯)，但有些句子較長，鑲嵌許多的關係子句，且加上修飾片語等等。例如在 (Graesser & Mcnamara, 2011) 的例子：

At anytime during the last 12 months, were you or any member of your

³ WordNet 的上下位關係是詞義與詞義 (sense) 間的關係，而非詞條 (lemma) 間的關係。此處僅是上下位關係的舉例說明，故不額外區分詞條以及其所包含的詞義。

household *enrolled* in or receiving benefits from free or reduced price meals at school through the Federal School Lunch program or the Federal School Breakfast program?

這段話有很長的名詞片語和修飾語；讀者需閱讀多個詞後才能遇到主要動詞 (*enrolled*)，造成較大的工作記憶負擔；句子中也有許多個邏輯連接詞 (*and, or*) 等。根據過去的研究指出，複雜的句法會讓讀者增加建立文本表徵時的工作記憶負擔 (Perfetti, Landi, & Oakhill, 2005)。Coh-Metrix 首先透過詞類標記器和句法剖析器幫助賦予每個句子句法樹 (*syntax tree*) 結構。並以句長、名詞片語的修飾詞數、主要動詞前的詞數來評估句法複雜度。同時，Coh-Metrix 也計算句子的句法相似度，其計算方法是根據句法樹的訊息，以兩句句法樹中相同的節點數做為分子，除以兩句中所有的節點數減去相同節點數，並以此比例作為兩句句法的相似度分數。過去研究指出，當文本中的句子共享類似地句法結構時，有助於讀者的閱讀理解 (Crossley, Greenfield, & McNamara, 2008)

以句法樹上的結構屬性來操作句法複雜度並不是唯一種操作句法複雜度的想法。句法樹的結構複雜度，背後的想法與詞彙與規則的處理機制取向較一致 (Pinker & Ullman, 2002)。該取向假設一組儲存詞彙知識的心理辭典，以及另一組抽象的形式 (文法) 規則，這些規則可用於合併、組合、轉換心理辭典的詞彙，這些文本中的語句需要應用較多的規則，則這個句子則較「複雜」。然而，這種看待句法複雜度的想法並非唯一的理論取徑。另一群研究者則認為語言使用是動態的 (Arnon & Snider, 2010)：讀者如何表徵、處理語言受到經驗、日常使用的影響。較常使用的語言單位會纏固 (*entrench*) 進讀者的記憶表徵，亦即文法規則的表徵不是靜態的，而是透過語言使用不斷形塑改變。各種語言粒度 (*granularity*) 的出現頻率則是不同語言單位在真實語言環境下的使用記錄。這些模式被統稱為「突顯模式」(*emergentist models*)，如語言使用取徑的文法理論 (Langacker, 1987)、連結論的語言理論 (Rumelhart & McClelland, 1986)、組塊模式 (Bransford & Franks, 1971)、或語言知識的事例模式

(exemplar models) (Gaul & Yu, 2006)。

過去的研究也的確發現，各種不同粒度的語言單位之出現頻率的確都會影響閱讀時的心理歷程。例如，在詞彙再認作業中，除了詞彙出現頻率（即詞頻效果）會影響作業表現外（Rayner & Duffy, 1986），多詞彙單位（multi-word phrases）的出現頻率（Arnon & Snider, 2010; Hsieh et al., 2017），甚至詞彙在句子中句法中的關係，例如詞彙在特定語法結構裡的頻率（Garnsey, Pearlmutter, Myers & Lotocky, 1997）、詞彙與特定論元的共現性（Trueswell, Tanenhaus, & Gurnsey, 1994）、整體的句法結構（Janssen & Barber, 2012）等都會影響讀者的表現。這些研究暗示，文本各階層的頻率訊息都會影響到讀者閱讀理解時的處理歷程（Diessel, 2007）。在適讀性研究中，大粒度（比詞彙單位大）之語言單位的出現頻率也是較少包括在文本屬性中的。

三、文本層次指標

文本層次指標脫離詞彙與句法等語言層次的屬性，進一步衡量與文本內容有關的工具。在這個層次中，讀者的目的是建立內容相關的一致性的文本表徵，Coh-Metrix 即衡量文本間是否具備足夠的指涉（reference）線索，讓讀者可以連接文本中不同的命題。

Coh-Metrix 從三個面向衡量指涉線索。第一是共同指涉（co-reference）指標。此項指標是評估文本中有多少的重複詞彙，可以幫助讀者建立命題連結。Coh-Metrix 有不同的方式定義詞彙重複度，例如當兩個句子彼此間有一個名詞是相同的，則這一對名詞則被定義為名詞重複。若兩個句子的兩個名詞的詞幹（stem）相同（例如 table/tables），則他們是論元重複。若兩個句子中，有一個句子的名詞（如，work）和另一個句子中任意詞類的詞（如，working）重複，則定義為詞幹重複。

另一種衡量文本指涉的方法，則是考量文章中的詞彙豐富度，亦即文本中用到多少不同的詞彙：如果文章中用到愈多不同的詞彙，讀者就愈難從單純地

詞彙重複找到連結命題的線索。Coh-Metrix 是以詞類/詞例的比 (type-token ratio) 來計算詞彙豐富度，當該比例愈高，代表文章中用了愈多不同的詞彙，讀者需要花費較多的心力將不同的詞彙整合進文本表徵。



第三個衡量文本表徵的則是用語意空間來計算概念的重複度。Coh-Metrix 使用潛在語意分析 (latent semantic analysis, LSA) (Landauer & Dumais, 1997) 從原始語料建構出語意空間，並以此代表讀者的世界知識。在 LSA 的模式中，每個詞彙都可以以一個多維向量代表，向量與向量之間的相似度則代表兩個詞彙的連結程度。過去研究亦指出，以 LSA 計算的詞彙連結度和人類參與者的自由聯想反應有顯著相關 (Landauer & Dumais, 1997)。Coh-Metrix 藉著不同詞彙在語意空間中的向量相似度，當作不同詞彙間的概念重複度。Coh-Metrix 也以句子作為單位，計算相鄰兩個句子間的概念重複度，或是文本中的特定句子與之前所有句子 (given-new) 的概念相似度 (McCarthy et al., 2012)。

四、情境模式指標

情境模式指標衡量文本是否有足夠的線索，讓讀者建立出文本所要敘述的事件或說明的內容。Zwaan 和 Radvansky (1998) 將情境模式分成五個面向：因果、目標、時間、空間和主角。若這五個面向中任意一個元素在文本中的關係不連貫，讀者所建立的情境模式將會受到阻礙。此時，則需要文本提供額外的連接詞、轉折語、副詞或其他線索，來幫助讀者察覺到訊息的不連貫，並幫助讀者連結前後文的訊息 (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007)。Coh-Metrix 將這些幫助建立情境模式的文本線索稱為助詞 (particles)，並將他們分為因果、意向和時間三類。例如，因果連接詞 (因為、使得)，時間副詞 (之前、之後)、意向詞 (為了，才能) 等等。Coh-Metrix 除了計算這些助詞出現的次數，同時也計算這些助詞 (如因果性助詞 because) 和相關目標動詞 (例如因果性動詞，move) 出現的比例，藉此評量文本在情境模式層次上提供的因果線索。

五、文本類型指標

文本類型指標試圖評估不同文本類型對於閱讀理解的影響。過去研究發現，即便內容領域相同、詞彙熟悉度相同，記敘文還是比論說文容易被讀者理解 (Graesser & McNamara, 2011)。讀者若能對不同類型的文章應用不同的閱讀技巧，則可增加其對文章的理解程度 (Meyer et al., 2010)。然而，一篇文章是記敘文或說明文，不僅在概念上並無一個清楚明確的劃界，在文本中也難以找到一組可操作的屬性來代表文本的類型。所以，這部份屬性雖然重要，但仍屬於需專業人力判斷及提供標記的領域。故文本類型並不是目前 Coh-Metrix 能自動產生指標的範圍。不過，在 Graesser 等人 (2011) 的主因子分析法中，仍發現有些表層的屬性可能與文章類型有關，例如動詞個數、代名詞個數、內容詞數目、熟悉度、負面詞彙等。

McNamara, Graesser, McCarthy 和 Cai (2014) 整合 Graesser 等人 (2011) 提出的主因子，然後再整理出一些其它的文本屬性，提出 Coh-Metrix 3.0 版。此版本的 Coh-Metrix 共包含 106 項指標，共 11 個類別；這些類別分別是描述性、主因子分數、指涉一致性、潛在語意分析、詞彙多樣性、連接詞數、情境模式、句法複雜度、句法樣式複雜度、詞彙訊息以及適讀性公式。其中主因子分數即來自於 Graesser 等人 (2011) 提出的 8 項主因子，他們分別對應到讀者閱讀時所建立不同層次的心理表徵，包括「詞彙豐富度因子」對應到詞彙層次、「句法簡單性因子」對應到句法層次、「指涉一致性因子」對應到文本表徵層次、「因果關係因子」、「動詞因子」、「邏輯因子」、「時間順序因子」對應到情境模式層次，以及「敘事因子」對應到文本類型層次。這 8 項因子共解釋 67.3% 的變異量，其中前 5 項主要因子 (敘事性、句法簡單性、詞彙具體性、指涉一致性和因果一致性) 已佔據 54% 的變異量。這個結果顯示雖然 Coh-Metrix 有 100 多個文本屬性，但背後有簡單的結構，有助於研究者用文本一致性的理論工具來分析文本，瞭解讀者閱讀時的心理歷程，甚至文本的適讀性。

肆、適讀性與文本一致性指標



Coh-Metrix 雖然臚列多項文本指標，但其本身並未定義哪些指標和文本適讀性有關。McNamara 等人 (2014) 認為，Coh-Metrix 的眾多指標適合以多向度的方式，以多個角度同時評量文本在閱讀時可能造成的困難。而且，Coh-Metrix 也能夠補足單向度的適讀性指標在衡量文本適讀性所造成的問題。例如，傳統的適讀性模式偏重於詞彙等文本表層屬性來決定文本的適讀性，這些模式未真正在屬性上考量到讀者在閱讀時所經歷的複雜認知歷程。其次，閱讀時，讀者會同時受到很多因素的影響，而傳統單一向度的適讀性模式難以考量到各種文本一致性屬性，以及敘述、說明文類等因素。最後，對實際教學者而言，他們需要的是知道學生在閱讀文本時遇到哪些困難，以及如何幫助學生提升閱讀表現 (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007)。亦即，理想的適讀性評量應該是診斷的過程，而這些訊息難以從單向度的適讀性公式獲得。

文獻中也的確指出 Coh-Metrix 有助於預測文章的適讀性。Dufty、Graesser、Louwrese 及 McNamara (2006) 即進行了 Coh-Metrix 和文章適讀性的效度研究。他們蒐集了 311 篇讀物，每本讀物都有出版社依適讀性公式及專家評定所設定的適讀年級。該研究以這些讀物的適讀年級為效標，結果發現適讀年級與 Flesch-Kincaid Grade level 的相關是 0.77，和 LSA 一致性 (句子與文本) 的相關是 -0.53。若以適讀性公式和一致性屬性兩者共同預測適讀年級時，其判定係數為 0.68。此結果顯示，對文本難度的預測，Coh-Metrix 所測量的文本屬性比 LSA 好。雖然加入一致性屬性後，略微減低適讀性公式的效度，但因為一致性屬性具有理論上的意義，所以整體而言，這個取向的適讀性研究是有意義的。Nelson、Perfetti、D. Liben 及 M. Liben (2012) 的研究也得到相似的結果。他們發現 Coh-Metrix 的因子分數和文章的適讀年級以及學生的閱讀表現有

顯著相關。其中句法簡易性、指涉一致性和敘事性是最能預測適讀性和閱讀表現的文本屬性。

Sung、Chang、Lin、Hsieh 及 Chang (2016) 年也發展適合計算中文文本適讀性的適讀性指標探索器 (Chinese Readability Index Explorer, CRIE)。該系統計算 70 個指標，共包含詞彙、句法、語義和篇章一致性四個層次。在詞彙層次下有五個子類別，分別是漢字複雜度、詞長、字詞數、頻率與詞彙豐富度。句法層次下則以句數、平均句長、修飾詞數等衡量句法難度。語義層次則以內容詞數、領域詞數、功能詞實詞比例等屬性評估語義複雜度。篇章一致性則使用指涉詞 (如代名詞) 數，連接詞數等操作文章一致性。除了計算指標之外，研究者更進一步將這些指標具體應用於適讀性預測上。Sung 等人 (2015) 以 CRIE 的指標作為文本屬性，並應用傳統線性模式和支持向量機 (support vector machine)，預測小學一到六年級的國語課文的所在年級。他們發現傳統線性模式和支持向量機分別最高可達 58.79%、71.75% 的預測正確率。

文本一致性指標 (或 Coh-Metrix 指標) 和過去傳統適讀性公式都同樣用文本屬性，來描述文章的適讀性。但和過去非常不同的是，文本一致性指標援引許多深層的文本屬性 (例如，句法剖析或語意表徵等)，來指涉人類閱讀時的「認知歷程」。在過去，這些深層的文本屬性需由人工標記、分析，但近年來自然語言處理技術蓬勃發展後，漸漸變成電腦可以勝任的工作，且在特定作業與領域 (domain) 下，電腦對於文本屬性的預測正確率已臻人工標記的水準。

除了 Coh-Metrix 這一系列的研究外，傳統適讀性公式的研究者以認知理論為基礎，提升適讀性公式的建構效度。例如，Smith、Stenner、Horabin、及 Smith (1989) 認為閱讀理解和語義單位的熟悉度與句法結構複雜性有關，所以他的適讀性公式以詞頻用來指涉閱讀中的語義屬性，句長當作語法複雜度屬性。類似地想法也反應在 Chall 和 Dale 在 1995 年提出的新公式中，他們使用句長和詞彙熟悉度作為公式中的預測變項，並增修了 Dale 與 Chall (1948) 使用

的常用詞表。Dale-Chall 的新公式與舊公式都和 Bormuth (1969) 的填空測驗都有非常高的相關 ($r = 0.92$)。Chall 與 Dale (1995) 的公式如下：

其公式如下：

$$\text{Lexile} = 9.82247 \times \text{平均句長}(\log) - 2.14634 \times \text{平均詞頻}(\log) - 3.23274$$

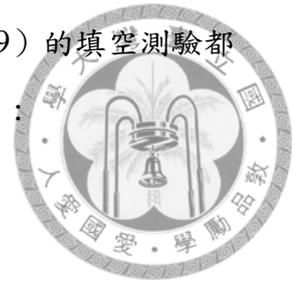
$$\text{New DaleChall} = 0.1579 \times \text{難詞百分比} + 0.0496 \times \text{平均句長}(\text{詞數})$$

Smith 等人 (1989) 以及 Chall 與 Dale (1995) 的研究顯示，傳統公式期的研究者並未在研究上排斥認知考量，反而，他們積極地試圖將能反應閱讀歷程的「認知變項」納入傳統適讀性公式。

第三節 文本特徵與自然語言處理

在過去 20 年間，電腦運算能力大幅成長，語言材料迅速累積，機器學習演算法蓬勃發展以及心理、語言學理論的進展，計算語言學以及自然語言處理領域顯著地成長 (Hirschberg & Manning, 2015)。尤其，近年來深度學習的發展更讓計算語言學中的各項作業表現 (如情緒極度分析、文章摘要、問答系統等) 大幅提升，傳統演算法和深度學習的類神經網絡如何共同幫助電腦理解人類的自然語言，仍是目前研究者熱切關心的議題 (Manning, 2015)。

在適讀性研究中，由於大量用到文本屬性 (包括表層語言屬性或文本一致性屬性)，為讓研究者有效率且客觀地計算操作文本，自然語言處理技術是相當適合的工程技術。自然語言處理在 1950 年間隨著當時對人工智慧的興趣開始發展，當時研究致力於發展一套讓電腦能掌握人類語言的「規則」。但研究者很快地發現人類語言的變異、複雜性和脈絡相依性讓電腦非常難以用「規則」來掌



握人類語言。在 1990 年間，自然語言處理開始使用機率模式 (statistical NLP) 和大量的標記語料，研究如何使用自然語料的統計特性，讓電腦學習標記和語言之間的關係，並用各種統計機率模式預測語言屬性，例如詞彙的詞類 (part-of-speech, POS)，情緒極度 (sentiment) 等屬性 (Hirschberg & Manning, 2015)。



自然語言處理有許多的次領域，例如從英文詞中找出詞幹 (stemming)、找出文章中的命名實體 (named entity recognition)、機器翻譯 (machine translation)、情感分析 (sentiment analysis)、文本摘要 (automatic summarization)、語音辨識 (speech recognition) 等。雖然上述應用目的眾多，各自都有深入細緻的演算法和模式，但大部份對於書面文本的自然語言處理都共享相對而言固定的基本處理。例如，詞的單位化 (tokenization)、詞類標記 (part-of-speech tagging)、句法剖析 (syntactic parsing) 等。這三項基本的處理如資料處理管線般之排列 (pipeline⁴)，原始文本從詞單位化開始，才能進行詞類標記，並藉由詞和詞類的訊息進行句法剖析 (Manning et al., 2014)。另外，當文本被單位化成詞彙後，自然語言處理即可藉助各種語意模式將詞彙的「語意」向量化成電腦容易處理的資料結構。這四項自然語言處理的過程在適讀性的文本一致性指標中扮演相當重要的角色。

壹、斷詞

在英文的自然語言處理中，第一個基礎的步驟是詞單位化 (tokenization)。在這個步驟中，文本裡的每個詞彙會依分析目的被單位化 (tokenize) 成一個個

⁴ 自然語言的處理管線 (pipeline) 是在技術上非常方便且常用的處理方式，但在心理語言學的文獻中，人類除了由下而上 (bottom-up) 的處理語言訊息外，也包含上到下 (top-down) 的處理歷程 (Greenberg, Healy, Koriati & Kreiner, 2004)。然而，的確有些研究者嘗試用處理管線以外的思維來幫助電腦掌握人類語言 (Ringgaard, Gupta, & Pereira, 2017)。

詞彙。在英文的書寫系統中，每個英文詞（word）通常會以空白分隔，所以大部分時候的單位化是用空白作為線索。只有在特定且少數的狀況下，英文仍會有跨空白的詞彙單位，例如「New York」或「rock ‘n’ roll」需要額外的處理，但大致上而言，詞單位化在英文是一個很清楚的處理作業。



然而，中文的詞單位化是很模糊且複雜的作業，由於其本質與英文的詞單位截然不同，故此作業被稱為中文「斷詞」（segmentation）（Ma & Chen, 2003；Hsieh, Tseng, Lee, & Chiang, 2018）。中文斷詞的困難在於中文書寫時以「字」為單位，但語言的概念表達是以詞彙（word）為單位。中文裡除了少數單音節詞是以一個「字」（character）表達外，多數是以二或多個音節（字）表達的複音詞。在文本缺乏其他線索（如空白）的情況下，電腦難以用簡單的符號線索區分出文本中包含哪些詞彙。而且，文本中常出現的斷詞歧異性，例如交集型歧異（「手機架」，可能是「手機」和「架」或「手」和「機架」），或組合型歧異（「才能」二字，在「他才能非凡」中應該是一個詞彙，但在「他才能勝任」，卻應該被斷成二個詞彙）。尤其，中文詞具備很高的流動性（Hoosain, 1992），各種新詞（小確幸）、專有名詞（谷歌）、仿真詞（factoid，如 11 月 30 日、23.5 度等）和構詞法衍生詞（如上下班、吃了飯）等都讓斷詞是一個非常具挑戰性的任務。

過去研究曾以兩種方法讓電腦自動分詞。首先，詞典法由研究者事先準備一組詞典，並以最大匹配法（又可細分為正向、逆向最大匹配和最少切分法）尋找文本裡面符合詞典裡的字串。若演算法發現有兩種以上可能的切分法，則搭配其他的經驗法則（heuristic）來解決歧異，例如，長詞優先、標準差較小者優先等。

第二種斷詞模式以統計模式搭配大量已人工事先斷詞的語料庫，讓電腦學習文本中的每個字是屬於詞首、詞中、詞尾或單獨成詞（Xue & Shen, 2003）。例如「台北市長明確表示」，統計模式法會將上述句子標記為「台(B)/北(E)/市

(B)/長(E)/明(B)/確(E)/表(B)/示(E)」，其中 B 代表該字是一個詞彙的詞首，E 是一個詞彙的詞尾。統計模式法應用不同的模式讓電腦學習字和類別的關係，例如最大熵值法 (maximum entropy) (Xue & Shen, 2003) 或條件隨機場 (conditional random field) 模式 (Tseng, Chang, Andrew, Jurafsky, & Manning, 2005) 等。統計模式法的優點在於遇到訓練材料未出現的新詞時 (out-of-vocabulary, OOV) 時，其斷詞正確率顯著高於詞典法 (黃昌寧、趙海，2007)。

在 2006 年的中文斷詞競賽中，各中文斷詞演算法的正確率可高達 97.2%，但有趣的是，若請人類標記者手動斷詞，則不同人之間的斷詞一致性平均為 76% (黃昌寧、趙海，2007)。胡志偉與方文熙 (1995) 在實驗中亦發現「斷詞歧異」的結果，他們認為「斷詞」是一項運用後設語言知識的行為，當參與者斷詞時，使用者會運用到「詞」以外的知識來進行作業，故會出現斷詞不一致的現象。

斷詞是中文自然語言處理的第一步，且在目前的自然語言處理技術中，斷詞也是之後詞頻計算、詞類標記、甚至句法剖析、建立語義空間等必須經過的步驟。但是，即便中文斷詞在演算法上已漸臻成熟，但中文「詞的地位」(wordhood) 的問題 (Huang, Hsieh, & Chen, 2017) 仍然深深影響著之後的每一步自然語言處理的結果和詮釋。

貳、詞類標記

在詞單位化或中文斷詞後，下一步則是瞭解每個詞的詞類 (part-of-speech)。英文的詞類有非常久遠的傳統，早在西元前 2 世紀，Thrax 的「語法的藝術」(The art of grammar) 就已經將詞彙分成 8 個類別：名詞、動詞、分詞 (participle)、冠詞 (article)、代名詞、介系詞、副詞和連接詞。這 8 個詞類直到現在並無太大的變動，僅有「分詞」的類別消失，並加入感嘆詞



(interjection)。相較之下，中文第一次引入西方語言的詞類觀念是在 19 世紀末的《馬氏文通》(馬建忠，1989)。該書中將漢語詞類分為九類，分別是五類實詞：名詞、代詞、動詞、形容詞和副詞，以及四類虛詞：介詞、連詞、助詞和嘆詞。晚近研究者如程祥徽與田小琳 (2015)，大致延續此分類架構，將漢語詞類分為 12 類：名詞、動詞、形容詞、數詞、量詞、代詞、副詞、介詞、連詞、助詞、歎詞和擬聲詞。亦有些研究者提出較不同的詞類定義，如中研院中文詞知識庫小組 (2004) 則以訊息為本的格位語法 (information-based case grammar, ICG) 界定出 8 類詞類，分別為述詞、非謂形容詞、體詞、副詞、介詞、連接詞、語助詞和感歎詞。

自然語言的詞類標記法大致上遵循著英文的自然語言研究發展。在英文的研究中，詞類標記可被視為是一個消歧 (disambiguation) 作業。亦即，雖然大多數的英文詞只有單一詞性 (80-86%)，但其他多詞性的詞卻有較高的使用頻率，在文本中約佔據了 55-67% 的詞例。詞類標記的目的是要賦予這些多詞類的詞彙正確的詞類。在詞類標記作業中，詞類通常來自於一個已標記好詞類的語料庫，該語料庫會定義一組詞類標籤集 (tagset)，讓電腦學習詞彙和詞類之間的關係。目前在漢語中常用的語料庫及詞類標籤集包含中央研究院詞庫小組的詞類 (中文詞知識庫小組，2004)，或者是中文句法庫 (Chinese Treebank) 所使用的賓州大學詞類集 (Chinese Penn Treebank part-of-speech tagset)。詞類標記的演算法常使用最大熵值法或隱式馬可夫模式 (Hidden Markov Model)

(Jurafsky & Martin, 2008)。其他研究亦嘗試使用雙向依存網路 (cyclic dependency network) 訓練詞類標記，該模式可在中文句法庫 (Chinese Treebank 7) 中達到 93.99% 的正確率 (Toutanova, Klein, Manning, & Singer, 2003)。

漢語的詞類標記的挑戰除了少了語言的屈折變化作為線索，漢語的特性更讓詞類標記的問題更具挑戰性。例如漢語的述詞本身 (有時再加上「的」或「地」之後) 就可以被用來修飾主要謂語。例如「他們非法入境」中的「非

法」本身是述詞，但在這個句子中卻是以「副詞」的形式修飾謂語「入境」。又例如，漢語的述詞常常可作名詞使用，像是「他努力完成研究」裡的「研究」本身是「述詞」，但在句子中卻被放在名詞的位置，甚至還可以被定量詞修飾，如「他努力完成那一項研究」。這些例子反映出當漢語詞彙被標記上詞類之後，其在句子中的功能可能還需要詞彙以外的線索來詮釋。



參、句法剖析

文本裡的句子不僅僅是一連串的词彙序列。句子中的詞彙是依循特定句法 (syntax) 規則組織成更大的單位 (如片語或句子)，甚至彼此包含或鑲嵌其中。在自然語言處理中，句法剖析的工作即試圖自動地將詞彙的句法關係具體的呈現出來。然而，由於自然語言的模糊性，即便一句簡單的句子，都可能有的句法剖析上的歧異性 (Jurafsky & Martin, 2008)。此歧異性可能來自語句的

(1) 附著模糊 (attachment ambiguity)。例如「他是我剛認識的朋友的小孩」，這句話中「剛認識的」修飾的可能是「朋友的小孩」，也有可能是「朋友」。或者是 (2) 並列模糊 (coordination ambiguity)，如「臺北和高雄的鄉下」，可以指「臺北和高雄」的鄉下，或者是「臺北」和「高雄的鄉下」。有些模糊性可仰賴統計上的出現機率協助區辨，亦即把較常出現的句法結構當成是最後的剖析結果。但是，有些句法模糊性需同時考量句法的統計特性、語意或脈絡知識才可能做出正確的判斷。亦即，句法剖析所產生的只是一種對於合理的語句剖析方式，並不是正確或唯一的剖析結果。

句法剖析的結果可透過各種演算法或模式產生，但如何表徵語句結構則牽涉到語言學的句法理論。在自然語言處理中，目前最常見的句法表徵方式包括成分語法 (constituency grammar) 和依存語法 (dependency grammar)。成分語法和 Chomsky (1956) 的片語結構 (phrase structure) 有密切關係，該理論認為



句子是由一個個「成分」組成的，每個成分可能包含一或多個詞彙，或甚至是其他成分。每個成分都有其衍生（derivation）規則，例如一個句子可能是由名詞片語（noun phrase, NP）和動詞片語（verb phrase, VP）構成；名詞片語的衍生規則可能又包含是由代名詞、專有名詞、定詞加上名詞等規則；動詞片語則可能包含動詞加名詞片語加上介系詞片語（preposition phrase, PP）等。成分的衍生關係通常會由一個句法樹所表達，從最基礎的根節點（root）代表句子中最大的成分，每個成分的衍生規則是句法樹的分支，直到衍生規則的成分是詞彙本身，或稱之為末端符號（terminal）為止。成分語法的剖析器可由動態規劃的演算法（如 Cocke-Younger-Kasami 演算法，簡稱 CYK 演算法；或 Earley 演算法）搭配上語言的衍生規則，有效率的找出一個句子的句法樹表徵。由於該衍生規則只和成分和其構成的元素有關，而與其他脈絡無關，故稱為脈絡獨立語法（context free grammar, CFG）。若按照該組衍生規則剖析句子時，出現一個以上的可能句法表徵，則可進一步透過語言規則在語料庫出現的機率，計算「機率性脈絡獨立語法」（probabilistic context free grammar, PCFG）計算最有可能出現的句法樹當作「正確」的句法表徵（Jurafsky & Martin, 2008）。

相較於「成分語法」重視句子由哪些成分（動詞片語、名詞片語等）組成，「依存語法」則強調詞彙與詞彙的互相依存關係。在「成分語法」中，當成分被剖析出來後，研究者通常會感興趣成分中的「中心語」以及相依項（dependent），但這些訊息需要透過其他的規則才能找出來。「依存語法」則直接試圖表徵句子中每個詞彙之間的「中心語」與「依存」的關係。這個特色讓依存語法和句子中的詞彙語義關係更明顯（Covington, 2001），例如 Kintsch 與 Mangalath（2011）就指出依存關係可直接地反應出文本的述詞及論元，更適合描述文本表徵中的命題結構。甚至，近年來依存語法剖析器不僅可分析詞彙與詞彙的相依關係，甚至可分析出該相依關係的類型（Nivre et al., 2016）：例如，「我吃飯」在依存語法裡會被剖析成「吃」是主要的詞彙，而「我」依存於



「吃」，是「主詞」(NSUBJ) 關係；「飯」也依存於「吃」，是「直接受詞」關係 (DOBJ)。依存語法在剖析時使用的是推移消滅剖析法 (shift-reduced parsing)。它在剖析時將一個個詞彙依序輸入緩衝區，剖析器會以機器學習模式決定目前的輸入詞彙，是否為緩衝區中詞彙的中心語或是依存項；如果一個詞彙已經找到其依存的中心語，則該詞彙將移出緩衝區。如果兩者無法建立任何關係，則將輸入詞彙放入緩衝區，再繼續嘗試下一個詞彙。在近年來，剖析器用來決定依存關係的方法可以是研究者手動撰寫的規則，當然也可能使用機器學習模式，例如是最大熵值模式、支持向量機或者是近年來盛行的深度學習等 (Chen & Manning, 2014; Jurafsky & Martin, 2008)。

肆、語意向量空間

文本的語義可分為詞彙語義 (lexical semantics) 和結構語義 (structural semantics)。前者指每個詞彙單獨所指涉的意義，後者則是由詞彙組成的結構所決定的意思 (Kearns, 2011)。例如「貓追老鼠」和「老鼠追貓」包含完全相同的詞彙 (貓、追、老鼠)，但這兩者由於詞彙的組成結構 (即詞序) 不同，兩者的結構訊息是截然不同的意思。這兩個層次的語義訊息在文本閱讀或適讀性研究中都扮演相當重要的角色，但此處主要討論詞彙語義的層次。

在適讀性研究中，研究者通常想藉由詞彙語義分析文本中出現的詞彙彼此之間的相似程度，藉此推論文本的詞彙概念上的一致性。在自然語言處理中，計算詞彙語義相似度有二種可能的取向 (Jurafsky & Martin, 2008)。首先是透過同義詞林或相關的語言資源 (language resources) 分析詞彙在分類上的相似程度。例如詞彙網路 (WordNet) (Fellbaum, 1998) 不僅提供詞彙同義的訊息，更提供每個意義之間的關係，可用來當作詞彙相似度的計算來源。

另一個取向則是以每個詞在語料中出現的環境當作線索，用共同出現的詞



彼此描述詞彙在語義空間中的位置。這類語義空間的操作取徑，源自於 Firth (1957) 所提出的「你可從詞之共現環境了解詞彙」(You shall know a word by the company it keeps)。亦即，詞彙的意義就是和它一起出現的詞彙分布。如果這些詞彙分布都可表徵在一個共同的空間中，則該空間就是詞彙的語義空間。然而在語料中，每個詞彙都與會非常多不同的詞彙共同出現，但大多數的共現詞彙次數又非常少。為了解決此稀缺性 (sparsity) 的問題，自然語言處理中使用降低維度 (dimension reduction) 的方法，將一個詞彙在語義空間的向量轉變為一組緊密向量 (dense vector)。該向量的優點是在比較不同詞彙語意時，更能掌握詞彙之間的語意相似度。

潛在語義分析 (latent semantic analysis, LSA) (Landauer & Dumais, 1997) 是其中一種降低維度的方法，也是在過去適讀性文獻中最常見的語義向量空間之建構方式。潛在語義分析是把語料中的詞彙以及每個詞彙所出現的文章分析成詞彙文章矩陣。該文章矩陣可能包含相當大的維度，但大多數元素都是 0 (因為大部分詞彙都沒有出現在大部分文章中)。潛在語義分析的方法是用歧異值分解 (singular value decomposition) 將該詞彙文章矩陣降低維度，讓每個詞彙或文章的向量都投影到一個較低維度的空間。詞彙與詞彙即可透過向量與向量之間的相似度計算，如餘弦相似度 (cosine similarity)，來決定彼此的相似度。在 Landauer 與 Dumais (1997) 的實驗中，他們對一個包含 30,473 篇文章，共 460 萬詞的語料庫中進行潛在語義分析，並進行如 TOEFL 測驗中的同義詞測驗。他們發現當潛在維度為 300 的時候，同義詞判斷作業的正確率最高 (51.5%)，且已到達英語為第二外語的考生水準。另一個降低語義空間維度的方法，是訓練神經網路模式讓機器從一個詞彙預測其周邊詞彙 (skip-gram)，並用過程中學習到的詞彙表徵當作它在語義空間中的表徵，例如 word2vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013)。Word2vec 和 LSA 相比，其詞彙表徵形式幾乎相同，其優點是訓練速度較快且較有效率 (Jurafsky &

Martin, 2008)，這在語料日益龐大的現代是一個很重要的特色。

另一種建構語義空間的方法是主題模式 (topic model) (Griffiths, Steyvers, & Tenenbaum, 2007)，它是根據 Blei 等人 (2003) 提出的隱式狄氏分配 (latent Dirichlet allocation) 加以延伸，並用來解釋並預測人類詞彙連結的數學形式。

主題模式假設文章背後都有一組主題分配 (distribution of topics)，文章中每個詞的主題都是從該主題分配抽取出來的。每個詞在文章中出現的機率，就是該詞從主題的詞彙分配 (distribution of words) 中被抽取出的機率。在主題模式和潛在語義分析都可以被視為一種將原本的詞彙文章矩陣投影到低維度語義空間的模式，但由於主題模式對主題分配和詞彙分配的表徵形式是機率函數，亦即每個詞彙表徵的元素都必須大於或等於零而且每個元素的總和必須是 1。這些特性讓主題模式在預測詞彙連結時可以用條件機率的方式表達詞彙關係，而不需受限於詞彙在語義空間中的距離。Griffiths 等人 (2007) 的研究比較主題模式與潛在語義分析的結果，他們發現主題模式比潛在語義分析更準確預測人類在詞彙自由聯想中的反應。Kintsch 與 Mangalath (2011) 亦認為主題模式可以區辨同一詞彙在不同脈絡下的語義，故更適合用來計算讀者於閱讀時的意義表徵。

語義向量空間的建構模式使用到不同的演算法或不同的形式表徵，但其輸入來源，無論是潛在語義分析的詞彙文章矩陣，word2vec 的鄰近詞 (skip-gram) 或連續詞袋 (continuous bags of words)，或者是主題模式中各文章的詞彙次數分配，都仍預設詞袋 (bag of words) 足以代表語義訊息。亦即，模式只需掌握一個文章中出現哪些詞彙，詞彙在文本或句子中出現的位置或順序並不重要。從過去文獻來看，這些模式的確能在各個自然語言作業中有優異的表現，如語義關連度、同義詞偵測、概念歸類、類比作業 (Baroni, Dinu, & Kruszewski, 2014)，或詞彙自由聯想 (Griffiths et al., 2007) 等。但上述作業主要都是和詞彙語義的層次相關，文本中的篇章、句子層次的語義，由於關連到

詞彙語義之外的訊息，仍然非常難以用上述模式加以衡量與操作。

斷詞、詞類標記、句法剖析和語義表徵都是在自然語言處理中相當成熟的作業，各種機器學習的模式也讓這些作業的預測正確率日益精進。藉由這些模式產生出各類的文本屬性，的確讓適讀性具備更多深層訊息，來指涉讀者閱讀文本時心中的理解歷程。然而，當適讀性從「認知解釋」的角度提出眾多的文本屬性，這些文本屬性卻需透過一個個「計算預測」作業取得。而且，從上述的討論中，每個預測作業的高正確率背後，又隱含自然語言的模糊性以及不同語言理論的假設。除此之外，適讀性研究還需將數量龐大的文本屬性（如 Coh-Metrix 具有 100 多個文本屬性）彙整成一個對於文本理解難度的評估，且文本屬性和適讀性之間的關係可能相當複雜，此彙整的方法逐漸從過去常見的線性迴歸，採用較以預測為導向的機器學習模式（Schwarm & Ostendorf, 2005）。

第四節 診斷式的適讀性系統

在適讀性研究的發展中，早期研究使用表層文本屬性預測文本的適讀性（本章第一節）；在閱讀的認知研究逐漸累積成果，深化人們對閱讀理解歷程的認識（本章第二節），這時有研究者企圖將文本一致性指標納入適讀性模式中。近年來受惠於自然語言處理與機器學習的發展（本章第三節），研究者可抽取出更多樣的文本屬性，以開發出預測正確率較高的適讀性模式。這個趨勢似乎隱含以計算預測為考量的取向和以認知解釋為考量的取向有若單維向度或光譜上的兩個端點，而在不同時代中，適讀性的研究派典在這光譜中來回擺盪。然而，本研究認為「認知解釋」以及「計算預測」此二取向不僅是在單軸向上互相擱抗的兩種考量；而是研究者在另一脈絡的兩個軸向上做出選擇後，所形成的表象。該脈絡至少包含二軸度，此兩個軸度分別是「輸入透明度」以及「參數透明度」。

壹、輸入與參數透明度



此章節將回顧 3 種常用來建構小學文本適讀性模式的取向，分別是：線性迴歸 (linear regression)、支持向量機 (support vector machine, SVM)、深度學習模式 (deep learning)，以及一個和文本閱讀十分有關，但較少在適讀性文獻中提及的主題模式 (topic modeling)。

此 4 種取向都可以建構適讀性模式，然而這些取向對如何達成預測文本適讀性的做法大相徑庭。根據這四種適讀性模式所輸入的資料可看出它們的差異。在以線性迴歸模式與支持向量機建構適讀性模式時，研究者需先透過特徵工程 (feature engineering)，在文本中選擇出與文本難度相關的文本屬性，並以這些屬性作為適讀性模式的輸入特徵 (Domingos, 2012)。特徵的選擇需要研究者對適讀性研究的背景知識，例如，他需要從過去的理論研究或自己的經驗中，發掘哪些文本特徵可能和文本的適讀性有關，並能對這些特徵做出適當轉換 (例如統計出文本的詞頻中數、或計算詞類詞例比率等)，使它們能成為模式的輸入資料。因為特徵的選擇與轉換與過去研究者對適讀性的理論想法高度相關，所以每個被選出的特徵都有豐富的實徵研究與文獻支持它與文本難度或與閱讀認知歷程有關係，故本研究將此取向稱為「高輸入透明度」。

有些研究者則是認為特徵工程的準備過程太過勞力密集，而且可能存在一些難以用口語描述，或現有知識理論中沒有涉及的特徵，這些特徵可以提升適讀性模式的預測效度。所以，若特徵產生的過程也能由模式從原始資料 (如構成文本的詞彙) 自行學習，則可大大簡化準備特徵的過程，甚至提升模式預測結果 (Bengio, Courville & Vincent, 2013)。這類想法可概括稱為特徵學習 (feature learning) 或表徵學習 (representational learning)。以主題模式和深度學習模式建構適讀性模式時，對於輸入模式的文本資料就抱持著這樣的的想法。因為在這種建構適讀性模式的取向中，模式能自行進行表徵學習 (亦即，

能從文本所含詞彙學習到哪些特徵與文本的適讀性有關)，所以這些模式的輸入資料是原始的文本詞彙，而詞彙與文本適讀性的關係是由模式決定。在過去的適讀性研究與理論中，從沒有提過「文本所用的所有詞彙會影響文本的適讀性」，故本研究將之這種輸入模式的方式稱為「低輸入透明度」。



另一種看待這四個模式的方法則來自於 Brieman (2001) 的看法。Brieman 認為適讀性模式的背後，蘊含著兩類的「文化」(culture)。不同文化對於如何從資料 (x) 導出預測值 (y) 的想法截然不同。第一種為「資料模式文化」

(Data modeling culture)。認同這種文化的研究者認為，預測模式需透過描述資料與效標變項間的關係來達成預測。此關係可能是兩者之間在真實世界中的統計關係、或資料產生機制。在適讀性研究中，以線性迴歸或主題模式建構適讀性模式的取向可歸在此類別。另一種是「演算法模式文化」(Algorithmic modeling culture)。認同這種文化的研究者認為，x 與 y 之間的關係可能是一個相當複雜的函數，這個函數本身不必然需符合真實世界中的 x 與 y 的關係，只要能從 x 預測出 y 即可。在適讀性研究中，以支持向量機和深度學習建構適讀性模式的取向可歸於此類別。

資料模式文化與演算法模式文化這兩類模式，其明顯差異在於模式的架構與模式參數。線性迴歸和主題模式取向的模式架構與參數與資料模式文化一致，本研究這兩種建構適讀性模式的取向分類為「高參數透明度模式」。基於同樣的原因，本研究把與「演算法文化」一致的支持向量機和深度學習模式分類為「低參數透明度模式」。因為本研究將實際的建構出這四種取向的適讀性模式，所以本章節接下來將分別討論（一）這四種模式是如何將預測文本適讀性的問題形式化的，（二）在這四種模式中，其輸入資料和輸出預測的關係，以及（三）在形式化的過程中，模式中的參數如何扮演不同的角色。

一、線性迴歸模式

線性迴歸是過去適讀性研究中長久以來泛用的數學模式。在以線性迴歸建

構適讀性模式時，每個文本屬性都視作為獨立（independent）的解釋變項（X），當文本的適讀年級被當作被預測的效標變項（y）時，它被視作為連續的正整數。線性迴歸方程式可用矩陣方法表達如下（假設 X 已包含截距項）：

$$y = \mathbf{X}\beta + e = \hat{y} + e \quad (1.1)$$

假設資料中有 M 篇文本，則這些文本在課本的年級將構成在高維空間的一組向量（其向量長度為 M）。其中，解釋變項所構成的矩陣 X，其維度為 $M \times K$ （K 為解釋變項個數）。在適讀性模式中，解釋變項的個數需小於文本個數，故 X 在 M 維度空間中，是由 K 個基底向量所構成的子空間。預測適讀性在線性迴歸模式是將高維度的 m 向量投影至由解釋變項（文本屬性）所形成的低維度空間（例如，若研究中使用了 77 個文本屬性，則構成 77 個維度的向量空間）。此空間中的基底（bases）彼此不必然正交，但皆是獨立的基底。迴歸模式透過最小方差法（least squared）尋找高維度空間在此低維度空間的最大投影量，其迴歸係數為 β ，並由矩陣運算求得參數。

迴歸模式希望高維度空間（長度 M）的向量投影至解釋變項所構成的低維度空間的誤差要最小，亦即誤差向量（e）需正交於低維度空間（X）。藉由向量正交的內積等於 0 的屬性， β 參數可用以下方法估計：

$$\begin{aligned} e &= y - \hat{y} \\ \mathbf{X}^T \cdot e &= 0 \\ \mathbf{X}^T \cdot (y - \mathbf{X}\beta) &= 0 \\ \beta &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T y \end{aligned} \quad (1.2)$$

式子 1.2 顯示在迴歸模式中，迴歸係數是從矩陣運算中計算而得的。在預

測時，則以一篇文章的屬性乘上係數，得到預測的年級。由於效標變項是連續的，故從預測年級到實際的適讀年級間需要將預測值轉換到最近的整數，當作適讀年級預測。



在迴歸模式中，文本屬性與適讀年級的關係皆是矩陣運算，其特色是每個屬性都會對應到明確的係數，此數值具明確的解釋意義。例如，文本中用字的筆畫平均數在迴歸模式中對應的係數為 0.08，則代表平均而言，在其他變項皆不變的條件下，平均筆畫數多 1，則文本的適讀年級會上升 0.08。除了迴歸模式的截距項（若不對解釋變項做其他轉換處理的條件下）可能不具備直覺上的意義外，模式中的每個係數都具備上述可述說的表達方式。

由於一旦經由線性迴歸建構出適讀性模式後，模式架構和參數就固定了，故研究者多致力於尋找能表徵文本難度的文本屬性，將這些屬性納入適讀性的迴歸模式。例如，從早期適讀性公式採用的表層屬性，到受認知心理學影響的 Coh-Metrix 屬性，甚至本研究為了操作語法複雜度所計算的罕見語式頻率屬性等，都是在這個脈絡下提出的。各種文本屬性都只是表達文本難度的一條取徑。在線性迴歸模式裡，各種不同的屬性或許只分別指涉表層難度、句法難度、或詞彙語意等層面的訊息，但研究者希望這些變項經由線性組合後，盡可能逼近「適讀性」的概念。

二、支持向量機

支持向量機和線性迴歸模式都是用文本屬性當作輸入資料，故兩者都是對文本屬性所構成的高維度空間做運算。然而，相較於線性迴歸模式是將高維度的向量，經由矩陣運算投影至低維度子空間，支持向量機是透過文本事例（instances）當作支持向量（support vectors），並藉由這些支持向量的線性組合，計算出可將文本正確歸類的分類平面。

假設輸入模式的訓練資料共有 M 筆（亦即 M 篇文章），每筆資料分別是一

組 \vec{u}_i 向量，其中包括 K 個數字（每篇文章皆有 K 個文本屬性），每個數字皆代表一個文本屬性的數值；每筆資料亦包含 1 個範圍在 1 至 6 的正整數，代表該篇文章所在之課本年級。支持向量機試圖在這群文本中找到一組方法，盡可能正確地將文本分成這 6 個類別。在此多重分類（multiclass）的問題中，研究者通常會用「一對其他」（one-vs-rest）的方法，將 6 個多重類別的分類問題，轉變為 6 個二元分類問題。每個二元分類問題的本質，都是將 m 篇文章分成屬於（或不屬於）某特定年級（如 5 年級）之兩類別。如果這二元分類問題是線性可分的（linear separable），亦即此空間中存在一個超平面（hyperplane）可區分這兩類別的事例。

在支持向量機中，該分類超平面可由超平面的法向量（normal vector） \vec{w} ，和截距項 b 來代表。此超平面將空間中的向量分成兩群：所有在超平面「以上」的事例都會大於 0，所有在超平面「以下」的事例都會小於 0。

$$\begin{aligned}\vec{w} \cdot \vec{u} + b &\geq 0 && \vec{u} \text{ 在超平面「以上」} \\ \vec{w} \cdot \vec{u} + b &\leq 0 && \vec{u} \text{ 在超平面「以下」}\end{aligned}\tag{1.3}$$

但是，在分類作業中，能最佳區分兩類事例的分類平面可能有非常非常多組。在支持向量機中，該模式認為最能區分兩類事例類別關係的分類平面，是能在兩類事例間留下最多「邊界」（margin）的超平面。亦即，模式不僅希望所有在超平面「以上」的事例都大於 0，還希望能大於某個值（例如 1），該值代表分類平面與實際事例之間的邊界。式子 1.4 的 y_i 代表事例 \vec{x}_i 的類別，若該事例在超平面以上，則 $y_i = 1$ ；相對地，則 $y_i = -1$ 。把 y_i 一起放入式子 1.3，可將兩個式子合併為式子 1.4。在式子 1.4 中亦把不等式皆代換成等式，亦即 \vec{x}_i 是剛好在邊界「上」的事例。

$$y_i (\vec{w} \cdot \vec{x}_i + b) - 1 = 0$$

(1.4)

支持向量機希望能最大化超平面（一個類別的邊界最大化）與將兩類別間最小化。此邊界也就是在超平面以上的 \vec{x}_+ ，和在超平面以下 \vec{x}_- 兩事例的向量差。此兩向量間的差異投影至超平面的法向量後，即是此超平面與事例之間的邊界。也就是邊界的寬度為：

$$\begin{aligned} \text{width} &= (\vec{x}_+ - \vec{x}_-) \cdot \frac{\vec{w}}{\|\vec{w}\|} \\ &= (\vec{x}_+ \vec{w} - \vec{x}_- \vec{w}) \cdot \frac{1}{\|\vec{w}\|} \\ &= (1 - b) + (1 + b) = \frac{2}{\|\vec{w}\|} \end{aligned}$$

(1.5)

式子 1.4 和 1.5 的結果顯示，若支持向量機想盡可能最大化（maximize）邊界，且該邊界可正確分類所有事例，則模式需解決的問題是（1）最小化 $\|\vec{w}\|$ ，以及（2）每個事例都要能被正確分類。上述兩目的可應用 Lagrange multiplier 轉變為另一種較容易處理的形式如下：

$$L = \frac{1}{2} \|\vec{w}\|^2 - \sum \alpha_i [y_i (\vec{w} \cdot \vec{x}_i + b) - 1]$$

(1.6)

式子 1.6 中，研究者關心的是 \vec{w} 的極值（extremum），亦即當 L 對 \vec{w} 的偏微分為 0 時的形式。

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum \alpha_i \cdot y_i \cdot \vec{x}_i = 0$$



從上述結果中，超平面的法向量 (\vec{w}) 具有如下形式：

$$\vec{w} = \sum \alpha_i \cdot y_i \cdot \vec{x}_i \quad (1.8)$$

式子 1.8 顯現的是支持向量機的重要特性，也就是分類超平面的法向量 (\vec{w}) 完全是由事例 (\vec{x}_i) 及某個待估計的模式參數 (α_i) 的線性加總而得。在支持向量機中，將這些影響到超平面的事例 (\vec{x}_i) 稱為支持向量。

在真實分類作業中，兩類事例不必然是如前述是線性可分的。有些事例無法被線性超平面完整的區分出兩個類別，故勢必有「分類錯誤」的情況。所以在此狀況下，模式不僅需最大化邊界，同時也最小化歸類錯誤：

$$\frac{1}{2} \|\vec{w}\|^2 + C \sum_i \xi_i^2$$
$$\xi_i = \max(0, 1 - y_i(\vec{w} \cdot \vec{x}_i - b)) \quad (1.9)$$

式子 1.9 中的 C 是支持向量機的懲罰參數，它可控制分類錯誤和最大邊界之間的比重。若 C 愈大，模式犯錯的「懲罰」會愈重，所以模式為避免「犯錯」，其分類邊界會變得較窄；反之，C 愈小，犯錯對於模式影響較小，其邊界就會較大。

在支持向量機中，模式真正估計的參數 (α_i) 和線性迴歸所估計的係數有截然不同的意義。此模式的參數並未直接與輸入的屬性相關，而是以特定的權重組合支持向量 (事例)，讓這些支持向量決定分類超平面。係數本身的意義與輸入資料或與文本適讀性年級之間並不直接有關。



三、主題模式

主題模式 (topic model)，又稱為隱藏狄氏分配 (latent Dirichlet allocation) (Blei et al., 2003; Griffiths et al., 2007)，屬於一種非監督式學習 (unsupervised learning) 的群聚方法 (clustering)。其輸入資料是原始的文本語料，亦即這些語料不需其它額外的標記輔助，即可從中估計出三個主要參數：每個詞彙背後所對應的主題 (topic)、文章的主題分配和主題的詞彙分配。

主題模式透過其參數，描述語料中每個文本以及每個詞彙的發生機率。假設在主題模式的訓練語料中包含了 M 篇文章，且所有的詞彙為 \mathbf{W} ，亦即 \mathbf{W} 包含 M 篇文章中所有出現的詞，每篇文章分別以 \mathbf{w} 和下標數字代表，如 $\mathbf{w}_1, \dots, \mathbf{w}_j, \dots, \mathbf{w}_M$ 。主題模式假設每篇文章 (\mathbf{w}_j) 的機率彼此是獨立的，所有文章的機率可拆解成各自文章的機率乘積：

$$p(\mathbf{W}) = \prod_{j=1}^M p(\mathbf{w}_j) \quad (1.10)$$

主題模式假設每個詞彙的發生機率都與它所屬的「主題」(topic) 有關。主題在模式中是一組潛在變項，文本中包含的每個詞彙 (\mathbf{w}_j) 背後都各自包含其對應的主題 (topic assignment) (\mathbf{z}_j)。所以，要描述每個詞彙的機率，應要一併考慮每個詞彙及其所對應的主題潛藏變項。亦即，每篇文章中詞彙的發生機率，都是來自於詞彙與主題之聯合機率，並對主題變項累加後的結果。

$$p(\mathbf{w}_j) = \sum_{z_j} p(\mathbf{w}_j, z_j) = \sum_{z_j} p(\mathbf{w}_j | z_j) p(z_j) \quad (1.11)$$

式子 1.11 顯示要描述詞彙的發生機率需要用到兩個成分：(1) $p(\mathbf{w}_j | z_j)$ 代

表每個詞彙在某個主題中的出現機率。主題模式是以模式參數， Φ ，計算此機率函數。該參數描述模式中所有主題的詞彙分配，也就是若模式中有 K 個主題和 V 個不同詞彙，則該參數是一個 $K \times V$ 的矩陣。矩陣中的每一列 (Φ_z) 代表所有詞彙在給定主題下的機率分配。(2) $p(\mathbf{z}_j)$ 則代表在某特定文章 (\mathbf{w}_j) 中，每個詞彙所對應主題 (\mathbf{z}_j) 之出現比例。模式計算該機率函數的方式是透過模式參數， Θ 。該參數是一個 $M \times K$ 的矩陣，每一列 (Θ_j) 代表所有主題在給定文章 (\mathbf{w}_j) 中的出現機率。

在主題模式中，只要給定該文章的主題分配後，文章內每個詞彙主題及詞彙的生成機率都是獨立的，所以式子 1.11 的文章詞彙機率和主題機率可以分解成每個詞彙的機率相乘。式子 1.12 中的 $w_{(j,t)}$ 表示第 j 篇文章（該文章包含 N_j 個詞彙）的第 t 個詞彙；同樣地， $z_{(j,t)}$ 代表 $w_{(j,t)}$ 所對應的主題。

$$p(\mathbf{w}_j | \mathbf{z}_j) p(\mathbf{z}_j) = \prod_{t=1}^{N_j} p(w_{(j,t)} | z_{(j,t)}, \Phi) p(z_{(j,t)} | \Theta_j)$$

$$z_{(j,t)} | \Theta_j \sim \text{Categorical}(\Theta_j)$$

$$w_{(j,t)} | \Phi_{z_{(j,t)}} \sim \text{Categorical}(\Phi_{z_{(j,t)}})$$

(1.12)

在主題模式中，主題分配參數 (Θ_j) 和各主題的詞彙分配參數 (Φ_i) 也是需按照資料推論的隨機變數。在主題模式中為求參數推論方便，常假設這兩個分配各自來自於類別分配的共軛先驗機率 (conjugate prior) 狄氏分配 (Dirichlet distribution)，該兩個參數的分配 (超) 參數 (hyperparameter) 分別為 α 和 β 。

$$\Theta_j \sim \text{Dirichlet}(\alpha), \text{ for } j \in [1..M]$$

$$\Phi_i \sim \text{Dirichlet}(\beta), \text{ for } i \in [1..K]$$



整個主題模式對文章詞彙的生成想法，可表達如下：

$$\begin{aligned}
 & p(\mathbf{W}, \mathbf{Z}, \Theta, \Phi | \alpha, \beta) \\
 &= \prod_{i=1}^K p(\Phi_k | \beta) \times \prod_{j=1}^M p(\Theta_j | \alpha) \times \prod_{t=1}^{N_j} p(z_{(j,t)} | \Theta_j) p(w_{(j,t)} | \Phi_{z_{(j,t)}})
 \end{aligned} \tag{1.14}$$

式子 1.14 是指一組文章所包含的詞彙、主題、以及模式參數這四個隨機變數的聯合機率，可分解成四項來源的乘積：(1) 詞彙主題分配的先驗機率、和 (2) 文章詞彙分配的先驗機率，以及 (3) 主題的機率、和 (4) 詞彙的機率。在這個聯合機率中，研究者通常真正關心的是三個潛在變項的事後機率 (posteriors)，也就是詞彙主題 (\mathbf{Z})、文章的主題分配 (Θ) 和主題的詞彙分配 (Φ)。這個機率推論的分母牽涉到同時對 Θ 與 Φ 兩個機率分配的積分，使得此機率僅能用逼近的方式求得結果。在文獻中，主題模式多用變異推論法 (variational inference) (Blei, 2003)，或 Gibbs 抽樣法 (Gibbs sampling) (Griffiths et al., 2007) 求得各潛在變項的事後機率近似值。

$$p(\mathbf{Z}, \Theta, \Phi | \mathbf{W}, \alpha, \beta) = \frac{p(\mathbf{W}, \mathbf{Z}, \Theta, \Phi | \alpha, \beta)}{p(\mathbf{W} | \alpha, \beta)} = \frac{p(\mathbf{W}, \mathbf{Z}, \Theta, \Phi | \alpha, \beta)}{\int_{\mathbf{Z}} \int_{\Theta} \int_{\Phi} p(\mathbf{W}, \mathbf{Z}, \Theta, \Phi | \alpha, \beta)} \tag{1.15}$$

上述三個潛在變項中，主題的詞彙分類 (Φ) 常被用來當作詞彙的語義空間模式，例如以詞彙分配當作語義向量的表徵，或用來預測詞彙的自由聯想 (Griffiths et al., 2007)。文章的主題分配 (Θ) 則常被應用於文本歸類作業 (document classification)，例如，一篇文章經由主題模式的估計後，該文章的主題分配是「0.8, 0.2」，該分配中的機率分別指主題一和主題二在文章中所佔的



比例。主題一的詞彙分配，可能包含「醫生、醫院、健保」等詞彙；主題二的機率分配可能包含「法案、政府、健保」等詞彙。從該主題所包含的詞彙中，研究者可賦予該主題一個標籤，例如主題一可能與醫療有關、主題二可能和政策有關。在這個例子中，由於主題一在該文章所佔的比例較大，因此可推知可能該文章與醫療有關。過去文獻中，研究者發現他們可用主題模式將不同期刊的文章分類到不同的學術領域，而且部份主題亦對應到特定的學術領域 (Blei, 2012)。這顯示主題模式的參數結果，具備分析文章類別和詞彙類別的特性。所以，主題模式雖然在過去的適讀性研究文獻中較少被提及，但它對於文本主題的參數估計結果很適合應用於文本適讀性所探討的問題。

四、深度學習模式

在本研究中，深度學習模式泛指各種應用類神經網絡 (artificial neural network) 來學習文本與適讀年級之間關係的監督式 (supervised) 機器學習模式。從機率的角度而言，深度學習模式所學習的仍是給定文本資料後，各個文本適讀年級的事後機率 (Ney, 1995)。只是，深度學習的模式中，估計此事後機率所用到的參數和模式架構並不直接與適讀性、文本、詞彙、甚至人類語言有直接關係。模式的目的是為讓它能更準確地從輸入資料間預測輸出結果 (文本所屬的年級)。此特性較接近 Brieman (2001) 所敘述的演算法模式文化。

近年來，深度學習模式提出眾多具備不同功能特性的網路層，但整體而言，模式都具備編碼 (encode) 和解碼 (decode) 的步驟 (Bahdanau, Cho, & Bengio, 2015; Cho et al., 2014; Devlin, Chang, Lee, & Toutanova, 2018; Mikolov et al., 2013)。在編碼步驟中，模式將文本所包含的詞彙序列轉變為一組模式可取用的向量表徵。此步驟又可分為兩個階段：首先，(1) 文本中所包含的每個詞彙，都先透過一個事先定義的詞彙對應表，對應到一個詞彙序號。例如，一篇有 N 個詞彙的文章，在轉變成詞彙序號之後，可以用一組長度為 N 的序號向量表達。這組序號向量 (每個序號都是一個正整數) 接著進入「嵌入層」

(embedding layer)。嵌入層包含一組 $V \times K$ 參數矩陣，其中 V 代表詞彙對應表的總詞數， K 則是嵌入層將以幾個數字代表一個詞彙。每個詞彙序號都直接對應到此參數矩陣中的一列數值。嵌入層的運算是將每個序號向量中的元素，都取代成在參數矩陣中對應列的數值。也就是在經過嵌入層後，文章的序號向量（維度為 N ），會轉變為一組二維的嵌入矩陣（ $N \times K$ ）。

(2) 此嵌入矩陣需經過進一步編碼 (encode) 成文本的模式表徵。此步驟有諸多可能的選擇：例如平均化層 (average pooling)、長短期記憶層 (long short-term memory, LSTM)、摺積層 (convolutional layer) 等。平均化層的特色是本身不具有其他參數，它僅是對嵌入矩陣的第一個維度做平均，使之轉變為一組維度為 K 的向量。其他的編碼網路層則需要各式參數，例如在 LSTM 中有許多組調控模式序列記憶的參數、摺積層則有摺積核心函數 (convolution kernels) 的大小和權重等參數等。

當模式將文本詞彙序列編碼成模式表徵後，解碼步驟則是從模式表徵中產生最後的適讀性預測。在此步驟中，模式表徵可進一步透過轉化層

(transformer) 操作模式表徵，或者直接以前饋網路 (feed forward) 連結至輸出層產生適讀性的預測。輸出層的節點數就是最後的適讀年級預測。若以 6 個適讀性年級為例，輸出層有 6 個節點，分別代表 1 至 6 年級。輸出節點的激發值將經過 softmax 函數轉換，讓適讀性的預測符合機率分配的性質。最後機率最高的節點就是模式的適讀性預測結果。此結果將和實際年級做比較，模式學習過程中，將盡可能調整模式中的所有參數，使得模式對實際年級的預測機率較高。在解碼步驟中，模式需要用到的參數包括從表徵到預測之間的網路層權重。例如以單層前饋網路連結模式表徵向量（維度為 K ）到輸出層，在前饋網路層中就需要一組維度為（ $6 \times K$ ）的參數矩陣。

在上述的編碼和解碼步驟中，深度學習模式至少會用到三群參數：(1) 建立嵌入矩陣的嵌入層參數；(2) 建立模式表徵的編碼網路層參數；(3) 解碼網

路層的參數。這些參數可提供深度學習模式很大的學習彈性，讓模式得以建立適當的模式表徵，甚至學習表徵與輸出之間的關係。然而，大部份的參數及模式表徵都是模式透過訓練資料自行學習，研究者難以事先知道模式參數與資料間的關係，大部份的參數也較難以用文本和適讀性的關係加以詮釋（Ribeiro, Singh, & Guestrin, 2016）。

貳、診斷式的適讀性

上述四種建構適讀性模式的取向（線性迴歸、支持向量機、主題模式與深度學習模式）各自具有不同的輸入透明度與參數透明度。根據 Shmueli（2010）對研究者以統計模式描述或解釋某種現象時，人們對該統計模式的直觀性理解會影響他們對模式推導出之現象的信任程度。如果我們接受 Shmueli 的論點，則上述四種取向所建構出的模式會因為它們在輸入與參數透明度上的不同，將影響在實際應用時，人們這些模式所提供之結果的信任。

演算法是否能讓人們理解其做出決定的過程，會影響人們對此系統的信任程度，特別是當此預測結果不符合人們的預期，或者是該預測結果和具體行動有關（Ribeiro et al., 2016）。Kizilcec（2016）在一項實徵研究中展現這個效果。該研究評估一個線上課程的評分系統。該系統會藉由一組統計方法校正學生互評的分數。在研究中，研究者藉操弄系統所提供的演算法說明來操作系統算則的透明度。研究結果發現，當使用者預期結果（學生自評）和系統提供的結果一致時，系統不需提供任何解釋，使用者就會對該系統有著最高的信任程度。但當使用者預期的結果和系統提供的結果不一致時，這時使用者會懷疑系統的結果，而讓使用者瞭解為什麼系統會得到該結果，可幫助使用者提升對該系統的信任程度。在適讀性模式的應用中，現在的機器學習發展加上適當的訓練資料，已經可以達到將近 8 成的正確率（Tseng, Hung, Sung, & Chen, 2016），這些

模式顯然已具有相當的應用價值。若使用者在運用適讀性工具時，不僅能了解模式的預測結果，也同時能了解造成文本難度較高的地方，甚至能得到如何修改文章難度的線索，應能提升適讀性模式的實際使用價值。



在本研究的設想中，一個診斷式的適讀性系統能運用不同取向的模式分析一篇文本的適讀性。這樣一來，它不僅能為使用者計算一篇文本的適讀性，它還能夠為使用者提供多項與文本適讀性有關的訊息。本研究希望這種作法能提升這個系統預測文本適讀性的透明度，從而使使用者接受系統對一篇文本適讀性的預測。甚至，使用者可以利用本系統提供的訊息編輯文章，改變文本的難度。我們可以用醫師對疾病的診斷比喻本系統的操作。醫師若要判斷病人是否得了感冒，他可能需要各種不同的訊息。例如，收集病人的生活作息，對身體狀況的主觀感受，對疾病的主觀研判等。當然，醫師也會收集病人的各種生理指標，如病人的體溫、心跳、呼吸的狀態、鼻咽喉是否紅腫、咳嗽、身體肌肉等反應。甚至醫師也會瞭解在目前環境中，是否有流行性感冒蔓延等。有些比較複雜的病症，醫生可能需用到更複雜的檢查，如血液中各項指數、X光、磁共振造影等不同的訊息來源。針對一位病患，醫生需綜合各種訊息，並輔以自身經驗，才能對病人的病況做出判斷。醫生的診斷過程並不是單純根據某些生理指標，經由某些線性組合、決定（超）平面（decision hyperplane）、或特徵之間的關係所決定，而是醫師、病人和訊息之間的相互作用的動態歷程。

本研究將分成幾個步驟建立診斷式的適讀性系統。首先，本研究將從模式的「輸入透明度」與「參數透明度」兩向度，區分出四種建構適讀性模式的取向，然後本研究將實際的建構出這些模式，並比較這些適讀性模式對文本適讀性預測的效度。其次，本研究以訪談和問卷法收集國小教師對不同模式的看法，以及他們是否會在實際教學場域中採用這些模式。最後本研究將參考教師對各種適讀性模式的看法，將系統中根據四種模式對文本適讀性評估的結果，選取教師們能夠理解的訊息，呈獻給他們參考。根據這些步驟，本論文將研究

將分為五個部分，依序呈現。它們分別是：文本屬性的抽取（研究一）、模式比較（研究二）、蒐集教師對模式信任程度的意見（研究三）、診斷性系統建置（研究四）以及實際的診斷系統評估（研究五）。



第二章 研究一、文本屬性分析



本研究使用國小一到六年級的課本，共有 950 篇課文，共 354,644 字。國語課文共有 439 篇⁵，186,956 字。社會（生活）課文共有 511 篇⁶，共 167,688 字。

本研究使用的國語課本包括。(一) 100 學年翰林版一到六年級的上學期課本、98 學年翰林版一到六年級的下學期課本、(二) 100 學年康軒版一到六年級的上學期課本、98 學年康軒版一到六年級的下學期課本 (三) 100 學年南一版一到六年級的上學期課本、99 學年南一版一到六年級的下學期課本。

本研究使用的社會課本包括：(一) 102 學年翰林版之一到二年級之生活課本，以及三到六年級之社會課本、(二) 102 學年康軒版一到二年級上學期的生活課本，及三到六年級的上學期的社會課本、101 學年康軒版一到二年級下學期的生活課本，及三到六年級下學期的社會課本、(三) 102 學年南一版一到二年級的上學期的生活課本、102 學年南一版三到六年級的社會課本，99 學年南一版一年級下學期的生活課本，以及 101 學年南一版二年級的生活課本，以及 101 學年三到六年級下學期的社會課本。

第一節 文本屬性的抽取與分析

研究一目的是按照過去傳統公式期和認知理論期的適讀性研究文獻 (Klare, 1976; Graesser et al., 2004)，找出與文本適讀性相關的文本屬性，並建立可自動

⁵ 本研究僅分析白話文的國語課文，故如文言文、新詩、劇本等文體皆不納入分析。

⁶ 生活科課本與社會科課本在九年一貫課程綱要中分別按照生活課程與社會學習課程編纂，嚴格而言並非屬同一課程。本研究為使社會領域文本與國語領域文本有相同類別數目之適讀年級，故將生活科與社會科文本合併納入社會領域文本。

抽取屬性的方式。按過去研究，本研究共選出 45 個過去在適讀性研究中相關的文本屬性，共 77 個指標（見附表 2.1）。這些指標亦可按其所分析的文本屬性分成六個類別：表層、詞彙、句法、語義、連接詞和一致性。



壹、研究方法

從文本抽取出屬性的過程分成三個階段：（一）文本前處理；（二）查找各項頻率常模、詞表或詞彙資料；（三）計算屬性數值。

文本前處理階段目的是將原始的文本變成計算屬性時所需用到的資料結構。此資料結構包含詞彙訊息以及句法樹結構。詞彙訊息包括文本經斷詞後所產生的各個詞彙，以及每個詞彙的詞類訊息（part of speech）。句法樹結構則是將文本中的每個句子（以句點、問號、驚嘆號分隔）轉變為代表其句法的樹狀結構。上述轉換過程牽涉到斷詞、詞類標記和句法剖析三項在自然語言處理中常用的作業。本研究選擇使用 Stanford CoreNLP（Manning et al., 2014）來完成上述作業。

Stanford CoreNLP 是 Stanford 大學的計算語言處理研究群所開發的程式，支援中文及其他多種語言。該程式可在該研究群的網頁上免費自由取得，且亦開放原始碼供各界參閱。該程式的斷詞模型是以隨機條件機率場判斷詞彙邊界，其斷詞正確率（F1-score）為.94（Chang, Galley & Manning, 2008）。詞類標記的模式則為循環依存網路（cyclic dependency network），詞類標記之正確率（F1-score）為.94（Toutanova et al., 2003）。在句法剖析的作業上，本研究同時使用該程式的成分剖析（constituency parsing）和依存剖析（dependency parsing）的結果。成分句法剖析是根據機率式脈絡獨立語法（PCFG）與 shift-reduce 剖析器找出句法結構，該模式的正確率（F1-score）為.86（Zhu, Zhang, Chen, Zhang & Zhu, 2013）。依存語法的句法剖析器是訓練神經網路模式預測每

個詞彙的依存關係，其正確率 (labeled attachment score, LAS) 為.82 (Chen & Manning, 2014)。



當文本經過前處理步驟後，文本中的詞彙僅包含詞類標記，本研究需進一步查找其他在計算屬性時用到的詞彙訊息。本研究使用到的詞彙訊息包括：

(1) 頻率訊息：由中研院平衡語料庫所計算出的字彙、詞彙的使用次數常模；
(2) 筆畫數訊息：從教育部國語辭典所取得的字筆畫訊息；(3) 分類詞表：包括動詞類別和連接詞類別的詞表，例如「或者」屬於選擇連接詞、「然後」屬於承接連接詞等。這些詞彙類別訊息來自於程祥微、田小琳 (2015) 的《現代漢語》；以及 (4) 詞彙語意數訊息：由中文詞彙網路 (黃居仁等人, 2008) 整理 7,960 個詞彙的語意數。在此步驟中，本研究透過程式自動查找上述詞彙訊息，供後續階段計算文本屬性使用。

分析的第三階段是計算文本屬性的確切數值。一篇文章經過前述處理後，即可按照詞彙訊息和句法樹結構計算本研究需使用的文本指標。因為指標的數目較多，為了討論上的方便，下文將文本屬性分為五類，以利說明指標的計算方法；這五類屬性分別為：表層類別、詞彙類別、句法類別、語義類別、一致性類別。

一、表層類別

表層類別的屬性指的是字筆畫數、詞長、子句長與句長四項僅需從文本表面即可計算數值的屬性。如字的筆畫數僅需要從字本身就可知道其筆畫，子句長只需以文中的逗點、冒號等其他分隔符號即可計算。雖然中文詞並非從文本表層可得知的訊息 (中文在書寫時無詞彙邊界訊息)，但為符合過去適讀性文獻的慣例，仍將詞長列為表層屬性。

這四項屬性分別屬於字、詞、子句、句長四個層次。在一篇文本中，可能包含多個字詞、子句或句子，故這些屬性在文本中將形成一個由多個數值組成的分配。本研究為盡可能保留這些屬性分配的訊息，故分別以三個統計指標描

述這個分配的第 1 四分位數、第 2 四分位數（中數），及第 3 四分位數。也就是說，從這四個屬性，本研究可以獲得 12 個與文本難度可能有關的指標。



二、詞彙類別

本研究整理出 21 個與詞彙有關的屬性：包括 1 個詞彙豐富度屬性（詞類詞例比）、2 個頻率屬性（字詞頻）、7 個詞彙比例屬性（名詞、形容詞等出現次數）、11 個連接詞類別屬性。詞彙豐富度屬性屬於文本層次的訊息，它計算的是文本中出現的詞例數目（word token）和不同詞類（word type）次數的比例。此指標的數值愈高，代表文本中用了愈多不同的詞。

文本中的字詞頻訊息來自於中研院平衡語料庫的字詞頻常模。在計算頻率屬性時，每個字例和詞例都會對應到一組使用次數。所以集合所有的字詞使用次數後，一篇文本中字和詞例的使用次數會分別形成一個分配。在文本中，字詞頻分布的使用頻率常常相當極端（Zipf, 1949）。此極端的分配型態常來自於文本中最常出現的詞彙通常都有特定句法角色。例如，英文中的功能詞（the、in、on 等）會在文本中多次的出現，所以這些詞在常模中的頻率也往往是高的；在中文中，文本中也有一些常用的詞彙（如，的、是、有、不、在等詞彙），這些詞彙在字詞頻常模的頻率也都相當高。然而，文本閱讀上的難度往往來自於其他使用頻率較低的字或詞。例如，Dale 與 Chall（1948）的適讀性模式中，就使用「難詞」的比率來預測文本的適讀性。這些在文本中各自扮演不同角色的詞彙，可能都蘊含與適讀性相關的訊息。本研究為捕捉文本中頻率分配的不同訊息，故使用兩種方法來描述字詞頻分布。一是以 5 個統計指標（第 1、第 5、第 25、第 50、第 75 百分位數）描述分配形狀。二則是將字詞按照使用次數的（常用）排序分成若干事先指定的範圍，並計算文本中有多少比例的字詞落在該範圍內。在字的層次中，本研究區分出 4 個範圍，分別是中研院平衡語料庫字頻常模排序前 799 的常用字、800（含）以上到 1500 以下、1500（含）到 3000，以及排序在 3000（含）以上的字彙。在詞的層次，則是根據中

研院平衡語料庫詞頻常模中排序在前 999 的常用詞，1000（含）到 2000、2000（含）到 5 萬、以及排序在 5 萬以後的詞彙。將字（詞）以其常用排序的位置劃分區段的優點是，字（詞）在各區段的分配比率較為平均，可改善頻率資料極端右偏的屬性。



詞類屬性包括分析各詞類在文本中出現的比例。此研究所計算的詞類比例共有名詞、動詞、形容詞、代名詞、連接詞、內容詞出現的比例。其中，動詞又區分為能願動詞（如能、要、會）與使令動詞（使、叫、讓）等二類。連接詞可提供句子之間的邏輯關係和時間的先後順序，對閱讀理解有很重要的功能。例如 Graesser 等人（2004）認為這些詞彙可以增加文本的一致性（text cohesion）。故本研究不但分析文本中連接詞的使用，更將其細分為並列、承接、遞進、選擇、轉折、因果、條件、假設、目的、解證 10 類（程祥徽、田小琳，2015），以便瞭解不同類別的連接詞與文本適讀性之間的關係。

一個詞彙的語義複雜度往往會影響讀者對這個詞彙辨識，以及含有這個詞彙之句子的處理（Rodd, Gaskell, & Marslen-Wilson, 2002; Foraker & Murphy, 2012）。在本研究中，一個詞彙之意義的複雜度根據的是該詞彙在中文詞彙網絡中的語意數目（黃居仁等人，2008）。在本研究中，一篇文本中的每個詞例都會對應到一個詞彙的語義數目，若一個詞彙未收錄於中文詞彙網絡中，本研究中將視之為單義詞。一篇文本中所用的所有詞例的語意數會形成一個分配，此分配的第 1、2、3 四分位數等三個統計指標將用以描述一篇文章所用詞例的語義複雜度分配。

三、句法類別

句子的句法複雜度會影響讀者對句子處理的時間，甚或理解的程度（Pearson, 2013），所以和句子有關的句法屬性也可能會影響文本的適讀性。本研究將透過 Stanford CoreNLP 對文本的句法剖析結果，將句子表徵為句法樹的形式，並參考 Coh-metrix 以及中文語法的特性，收集兩類與句法相關的屬性：

(1) 句法的結構；(2) 語式的頻率。

結構屬性。從一篇文本中，本研究不但可以計算類似整體文本長度（詞數）的表層屬性，還可以根據文本所含之句子的句法樹，計算出 6 項與片語（名詞片語、動詞片語、介係詞片語）及句子類型（把字句、被字句、比字句）相關的屬性的出現次數。把、被、比三種句式雖然在句子的表層結構中有明顯的單字詞（即「把」、「被」、「比」）作句型標記（marker），但這三個單字詞都可能會在句子中扮演其他角色。例如，「把」在「我/把/飯/吃/完/了」句子中的確是當作把字句的句法標記，但是同樣要做單字詞的「把」，在「我/撐/一/把/傘」中卻是當量詞用。故句式的分類仍需在對句子完成句法剖析後，才能分析句子類別。

名詞、動詞與介係詞片語之資料的計算則是透過 Stanford CoreNLP 中的成分語法的剖析結果。從分析的結果，本研究首先根據每個句子的句法樹，計算各片語類型的出現次數；其次，將一個片語類型的出現次數除以文本總詞數，這樣可以在控制文本長度的情況下，得到各片語類型在文本中出現的相對比例。

句法樹結構除了可以讓研究者計算上述與片語、句子類別的資料外，還可計算三種 Coh-Metrix 所涵蓋的句子結構的特性；它們分別是「主要動詞前詞彙數」（亦即，在一個句子中，位於該句主要動詞之前的詞彙數量；在 Coh-Metrix 中，它被稱為「左側嵌入」，left-embeddedness），名詞修飾詞數和句法相似性。本研究根據 Stanford CoreNLP 之依存語法的剖析結果計算前 2 項屬性。主要動詞前詞彙數的分析方式如下：（一）在一個句子的句法樹中找出這個句子直接依存於句法根元素（root）的動詞，（二）計算位於這個動詞前面的詞彙數目。例如，「他買了一本書」中的主要動詞是「買」，位於該動詞前面只有一個主詞「他」，故此句子的主要動詞前詞彙數為 1。一個名詞的修飾詞數是計算相依於該名詞的詞彙數目。例如，在「一本書」中，名詞「書」的相依詞為數詞

「一」與量詞「本」，故名詞書的修飾詞數是 2。如果相依詞的本身也有相依詞，則名詞之修飾詞的數目為所有相依詞的總和。例如，在「一整杯咖啡」中，「整」為「杯」的相依詞，所以「咖啡」的修飾詞數為 3。本研究將計算一個句子中所有名詞的飾詞數，然後以修飾詞數最大的名詞當作這個句子的指標。

兩個句子的句法相似性是比較它們的成分語法句法樹的相似性。確切的說，是計算兩個句法樹在相對應的位置上是否有同樣的句法樹節點。例如，在國小國語課文中有下面的句子：「一朵雲，兩朵雲，天上共有多少雲？」以及「一朵花，兩朵花，地上共有多少花？」這兩個句子的句法樹結構是相同的，所以兩句的句法相似度為即為 1。又例如，另外一句來自於國小國語課本的句子：「小鳥唱歌真好聽。樹葉說話細又輕」，此兩句子句型雖不完全相同，這整體而言仍相似，其句法相似度為 0.72。本研究將分析一篇文本中所有兩兩相鄰句子的句法相似性，並計算其平均作為文本的句法相似性屬性。

除了上述與 Coh-Metrix 相似的結構屬性外，本研究亦從句法樹結構中抽出句法深度這個可以與句法複雜度相關的屬性。本研究參考 Kintsch 與 Mangalath (2011) 的操作定義，根據 Stanford CoreNLP 之依存語法的剖析結果，計算之命題深度，以及成分語法的剖析結果計算之句法樹深度。根據 Kintsch 與 Mangalath (2011)，命題為一組相依存的詞彙。一個句子中，詞彙之間的依存關係可讓命題間，連結成一個命題的樹狀結構。句子的命題深度為這個命題樹的最大高度。以「地上的小花開了」這個句子為例，從圖 2.1 右圖中，我們可以計算出該句子的命題深度為 3。相似地，句法樹深度則是以成分句法剖析的結果，計算句子的句法樹高度。圖 2.1 左圖顯示「地上的小花開了」的句法樹。從句法樹的 ROOT 到最低的節點 NN (地上) 之間的深度為 6。

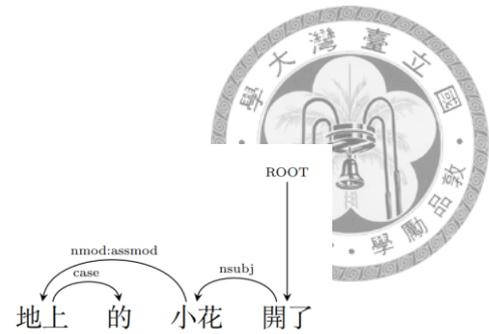
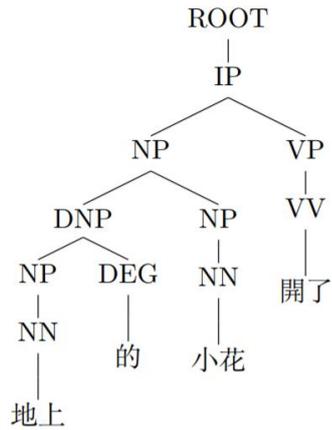


圖 2.1：句法剖析結果

註：左圖為成分句法剖析結果，右圖為依存句法剖析結果。原句取自小學國語一年級課文。

語式頻率屬性。本研究除了從句法樹結構中計算各種句子的句法特性外，亦試圖從結構中抽取出不同層次的「語式」，並分析這些語式包含的詞類序列。在本研究中，所謂的語式指的是句法樹中，包含兩個（或兩個以上）子節點的節點。以圖 2.1 的句子為例，我們可以對它做以下的分析：(1) 整句層次，亦即對 IP 節點的分析；它包含了 4 個詞（地上、的、小花、開了），這 4 個詞的詞類分別是 NN（名詞）、DEG（的）、NN（名詞）、VV（動詞）；(2) 針對 IP 之下的 NP 與 VP 的分析。例如，NP 節點下共包含 3 個詞（地上、的、小花），其詞類分別是 NN、DEG、NN；(3) 針對 NP 之下的節點進行分析。在本例句中，NP 之下有 DNP 及 NP 兩個結點。其中，DNP 節點包含 2 個詞（地上、的），其詞類分別是 NN、DEG。本研究不會繼續分析 DNP 以下的節點，因為這些節點之下只含有一個節點。

雖然一種語式出現在句法樹中的不同深度時，在語言的理解和產製上可能皆具備不同意義，但本研究考量到語式頻率的稀鬆性（sparsity）及簡化分析過程，故將出現在不同句法樹深度的語式視作相同的語式。上述分析的目的是，

計算各種語式在本研究的目標材料中（國小國語與社會課文）出現的頻率常模；本研究也將利用詞常模分析語式頻率與適讀性之間的關係。

為了建立國小國語及社會課文用語的語式頻率常模，本研究將所有研究材料的句法樹結構都分析成各種可能的語式，並計算這些語式的出現次數。從讀者對閱讀材料的熟悉度看，語式的出現次數和詞彙的出現詞頻具有非常類似地意義。亦即，若一種語式的出現次數愈低，代表這個語式的句法結構愈罕見，讀者對它也越不熟悉。本研究嘗試性的以此常模計算一個句子中出現的罕見語式的頻率，並希望探討這個屬性與適讀性之間的關係。從詞頻與適讀性的關係，我們可以合理的推論，如果一個句子包含幾個在常模中是僅出現 1 次的語式，這個句子的語法結構的「難度」應該是頗高的，而如果文本中包含大量這類的句子，文本的適讀性應該是低的。

以圖 2.1 的句子為例，該句子最上層的語式為 NN、DEG、NN、VV 在語式頻率常模中出現的次數為 9；第二層的語式為 NN、DEG、NN，頻率為 797；第三層的語式為 NN、DEG，頻率為 1087。故這個句子未包含任何只有出現 1 次的語式，其罕見語式頻率屬性為 0。

四、語義類別

本研究分析了幾種詞彙的語義特徵。第一種詞彙的語義特徵指的是詞彙的語義複雜度（亦即，語義數目）。本研究根據黃居仁等人（2008）公布的中文詞彙網路（包含之 7,960 個詞彙，25,657 個詞義）定義詞彙的語義複雜度。若一個詞未包括在詞彙網路中，其原因可能是該詞彙的頻率較低或為特定專有名詞，故在資料庫中未收錄。本分析為簡化分析步驟，故將這些未收錄的詞視作單義詞（詞義為 1）。

在本研究的主題模式取向的適讀性模式建構過程，研究者會分析國小國文及社會課文內含的主題，而藉著這些主題，本研究可以分析出兩種文本層次的語義指標。這兩種指標皆是藉著計算一個句子中的某個詞彙與其它文本中（出

現在它之前)的詞彙之間的連結機率 (word association) (Griffiths et al., 2007) (見式子 2.1), 分析這個詞彙與其它詞彙重複出現在各個主題中的機率。這種詞彙重複指標也和語義的複雜度有關。當一個句子中的所有詞彙的連結機率都計算出來之後, 這些機率的平均數將作為這個句子的語義重複性。



語意重複指標又可以按照文本中「其它詞」的定義範圍區分為語義局部重複指標和語義脈絡重複指標。在語義局部重複指標的定義中, 其它詞指的是, 目前分析之詞彙的前一個句子中所包含的詞彙, 而對語義脈絡重複指標而言, 其它詞指的是, 目前分析之詞彙的前文所出現的所有詞彙。

$$\begin{aligned}
 p(\mathbf{w}^{(y)}|\mathbf{w}^{(x)}) &= \sum_z p(\mathbf{w}^{(y)}, z|\mathbf{w}^{(x)}) \\
 &= \sum_z p(\mathbf{w}^{(y)}|z, \mathbf{w}^{(x)}) p(z|\mathbf{w}^{(x)}) \\
 &= \sum_z p(\mathbf{w}^{(y)}|z) \cdot \frac{p(\mathbf{w}^{(x)}|z)p(z)}{\sum_z p(\mathbf{w}^{(x)}|z)p(z)}
 \end{aligned}$$

$\mathbf{w}^{(x)}$ 、 $\mathbf{w}^{(y)}$ ：前句、與目前句所包含的各個詞彙

z ：詞彙所屬的主題(topic)

(2.1)

五、一致性類別

參考 Coh-matrix 的做法 (Graesser et al., 2004), 本研究兩類詞彙 (名詞與內容詞) 在文本中兩相鄰句子間的重複指標。例如, 「我今天不吃蘋果」和「他吃蘋果」是兩個在文本中相鄰的句子, 「吃蘋果」在兩個句子中重複的出現, 彰顯兩個句子間的一致性。其中, 「蘋果」是名詞, 他在兩個句子中重複的出現, 故計入名詞重複性屬性。「吃蘋果」這個片語則包含一個動詞與一個名詞, 因為

動詞與名詞均為內容詞，在 Coh-metrix 中，它們皆被計入「內容詞」重複性屬性。本研究分析了詞彙的局部重複指標，亦即目前正在分析的目標句與其前面的句子的一致性；而脈絡重複指標指的是，目前分析的句子與它前面所有句子之間的一致性。



貳、結果與討論

本研究以國小國語及社會（包括生活）課本為範疇，以這些課本中的 950 篇課文為分析的目標，計算前述 77 種文本指標。本研究也按照科目及年級將這些課文分為不同的文本組，分析各種文本指標。各文本指標的分析結果，將以年級中數的方式呈現於圖示於附圖 3.1-3.5。下文將簡略的討論本研究的分析結果。

一、語式頻率分析

在國語課文中，本研究總共分析出 43,651 種不同的語式，總次數為 116,568 次，其中只出現 1 次的語式有 36,244 個。最常出現的前五個語式是「一般動詞、名詞」（4,228 次）、「數詞、量詞」（2,923 次）、「名詞、名詞」（2,701 次），以及「一般動詞、一般動詞」（2,662 次）。前 100 個在國語課文中前 100 個在國語課文中最常出現的語式列於附表 1.1。在社會課文中，總共分析出 31,124 個語式，總次數為 84,601 次，其中只出現一次的語式有 25,330 次。最常出現的五個語式是「名詞、名詞」（4,484 次）、「一般動詞、名詞」（2,981 次）、「一般動詞、一般動詞」（1,300 次）、「一般動詞、名詞、名詞」（1,167 次）。前 100 個在社會課文中最常出現的語式列於附表 1.2。圖 2.2、圖 2.3 分別呈現國語課文和社會課文的語式長度（由幾個詞所構成的語式）和頻率的分布圖。二科目的頻率分布圖可顯現此分析方法的特徵，亦即絕大多數的頻率都只集中在少數短語式上，而大部份的長語式都僅出現一次。但從頻率分布

亦可看出語式長短並未和出現頻率完全相關，有些短語式的出現頻率非常低。

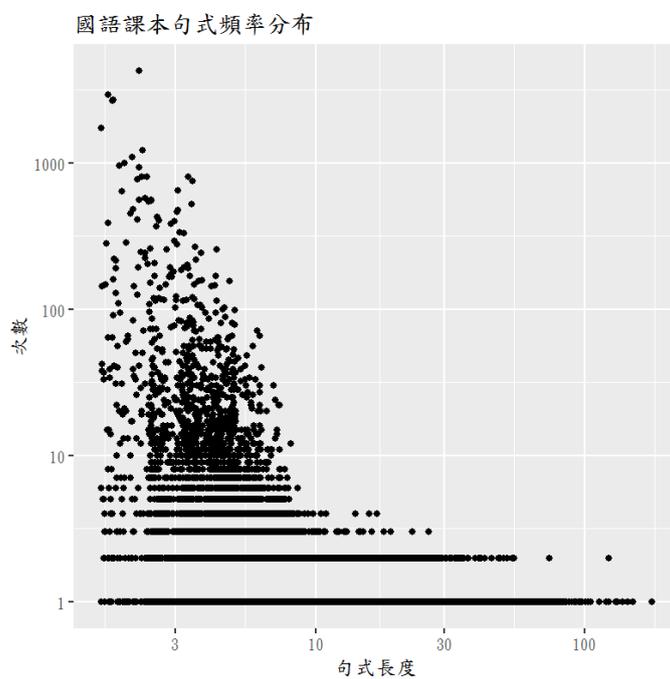


圖 2.2：國語課本語式頻率與長度分布

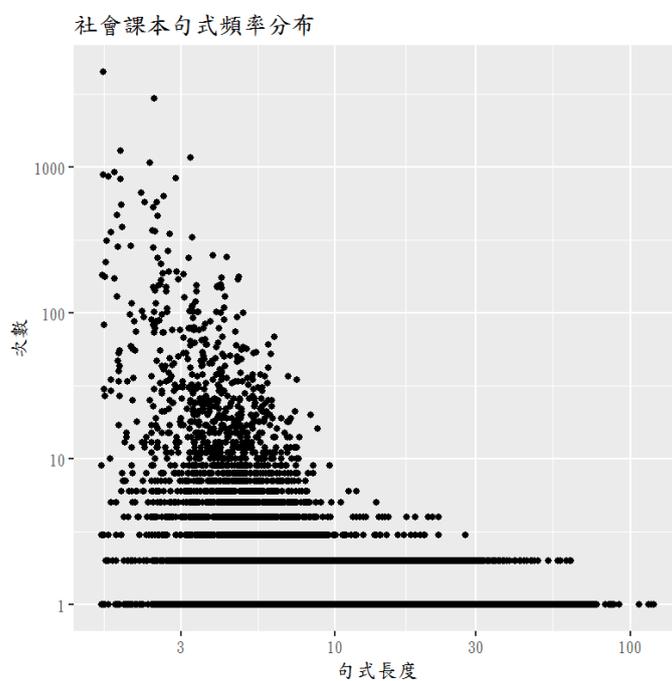


圖 2.3：社會課本語式頻率與長度分布

二、屬性結果與常模



理論而言，分析結果會形成這 77 個描述文本指標的常模。但由於為常模中所含資料過於龐大，所以除了部分已經發表的常模（曾昱翔、胡志偉、羅明、呂明蓁、呂菁菁，2014），以及指涉句法複雜度的部分常模，本論文將僅以圖的形式，呈現各常模的各年級中數（附圖 3.1-3.5）。從圖中可看出各種文本屬性與年級之間的關係。例如，整體而言，字頻、詞頻皆與年級間呈現負向趨勢，亦即愈高年級的文本，越可能出現較低頻率的字和詞（用較難的字與詞）。與年級間呈現類似負向相關的屬性還有子句長、代名詞比例、代名詞名詞比、詞類詞例比、動詞比例、動詞片語比例、詞彙語義數、內容詞功能詞比例、使令動詞比例、能願動詞比例、語義重複性指標、句法相似度等。和上述趨勢相反的是字筆畫數；它大致上和文本年級呈現正相關，亦即年級愈高的文本中，越可能出現多筆畫數的字。其它與年級有正向趨勢的文本屬性包括：句長、連接詞比例、名詞比例、名詞片語比例、介系詞片語比例、內容詞重複性指標、句法樹高度、句法樹最大深度、罕見片語語式頻率、名詞前修飾詞數、主要動詞前詞數、名詞重複性指標等。

在本文計算的 77 個文本指標中，與文本所在年級有最強正相關的前五個指標包括句長 ($r = .72$)、句法樹最大深度 ($r = .65$)、罕見片語語式頻率第 3 四分位數 ($r = .63$)、名詞修飾詞數 ($r = .60$)、句法樹高度第 3 四分位數 ($r = .57$)。這些指標很明顯的都和句法的複雜度相關。其中，句長與句法複雜度之間可能有著間接的關係；因為長句給予了複雜句法出現的機會。而其它四個指標則與句法複雜度有直接的關係。與文本所在年級有最強負相關的文本指標則是詞義數中數 ($r = -.51$)、代名詞比例 ($r = -.47$)、詞義數第 3 四分位數 ($r = -.46$)、動詞比例 ($r = -.45$)、以及代名詞與名詞比 ($r = -.44$)。在這些指標中，「代名詞比例」和「代名詞與名詞比」是早在 1930 年代，Flesch 用來評量文本適讀性的屬性。



本研究計算之 77 個文本指標來自於 45 個文本屬性，因為有些變項又被細分出幾種指標（例如，字頻被細分出第 1、5、25、50、75 百分位數五種指標）。也就是說，在本研究所分析的指標中，有些與傳統適讀性公式所含之屬性一樣，是「跨文本」的屬性，而有些指標（如字頻的幾種百分位數）則可以彰顯一篇文本內，某則可以彰顯一篇文本內，某種文本屬性的分布。

在跨文本指標方面，有些指標之間存在著相關。例如句法樹的最大深度和句長 ($r = .54$)、罕見片語語式次數 ($r = .52$)、名詞前修飾詞數 ($r = .50$) 間皆有顯著的高相關；這意味著這些指標間牽涉同一種潛在構念。這組高相關的指標可能均與文本中的句法複雜度有關。

有些指標來自於文本的同一屬性（屬性；例如，字頻的五種百分位數），而這些指標間的關係可以進一步讓研究者瞭解在一篇文本中，這個屬性的出現情形。研究結果發現，字頻的第 1 百分位數和第 5 百分位數間有顯著的高正相關 ($r = .81$)，這結果顯示，當文本出現低頻字（較難的字）時，也容易出現其它的低頻字。有趣的是，字頻的第 1 百分位數和第 75 百分位數 ($r = .16$) 的相關很低，這代表文本中的高頻字的使用與低頻字的使用無關。這個結果顯示，即便是一篇用字難度頗高的文本，文本中仍然會使用高頻字。

第二節 文本屬性與適讀年級迴歸分析

在傳統公式期，研究者分析各種文本屬性與適讀性之間的關係，並企圖建立適讀性模式來預測文本的閱讀難度或適讀年級。在這個期間，研究者最常用來建構適讀性模式的方法是，（一）收集一組已經標定難度（或適讀年級）的文本，（二）分析這些文本的各種可能與文本難讀有關的文本屬性，（三）以分析出的文本屬性為預測變項，文本難度（或適讀年級）為效標變項，進行統計迴歸分析。研究者根據統計分析的結果（亦即，迴歸公式），建構適讀性模式。



研究 1.2 將依循前述的適讀性模式建構方式，以研究 1.1 所計算出之各種文本屬性（77 種文本指標）為預測變項，文本所屬年即為效標變項，進行線性迴歸分析。此分析之目的為（1）重複驗證先前過去適讀性研究對於文本屬性與文本適讀年級的發現；（2）瞭解不同文本屬性在不同類型的文本中（國語與社會），是否對預測文本適讀性有不同的重要性。

本研究特別將 77 個文本指標分成三大類別：（1）表層屬性：包含詞長、子句長、句長、字筆畫數、字頻、詞頻以及其各項統計指標；（2）Coh-Metrix 屬性：在表層屬性之外，但不包括本研究所計算之罕見語式屬性以及句法樹深度之屬性。（3）完整屬性：包含所有 77 個文本屬性。其中，屬於表層類別的文本屬性多是傳統公式期研究所使用的屬性，而 Coh-Metrix 所包含的屬性則是在認知理論期，研究者提出並實際用以評估文本適讀性的文本屬性。最後，在完整屬性部分，本研究加入了幾個與句法複雜性有關的指標。也就是說，藉著本研究的分析，研究者可以了解，當一些認知心理學家認為會影響閱讀理解的文本屬性納入適讀性模式中時，是否會影響該模式對文本適讀性的預測。

一、研究方法

如前所述，本研究將 77 個文本指標分成三大類別：表層屬性、Coh-Metrix 屬性以及完整屬性。分析時，每個類別的指標都會被分別放入迴歸模式，並且以逐步迴歸法（stepwise regression）選擇對模式預測適讀年級有幫助的文本屬性。在逐步迴歸中，模式中的所有指標都先納入迴歸模式，在一次迭代中，模式試著移除每個指標，並計算該指標移除後對模式 AIC 指標的影響（T. Yamashita, K. Yamashita & Kamimura, 2007）。在該次迭代中，對 AIC 影響最小的變項就被移除（backward stepwise），並進入下一次迭代。變項選擇的程序將持續到移除指標不會再對模式 AIC 有幫助時停止。

為了了解不同類型文本（亦即，國語課文和社會課文）的是否會受到不同文本屬性的影響，本研究分別針對不同文本類型，獨立進行文本屬性選擇與適

讀性迴歸分析。故研究 1.2 總共計算 9 種適讀性迴歸模式。這 9 種模式代表的是在三種文本類型（全部課文、國語課文和社會課文）與三種迴歸模式（表層模式、Coh-Metrix 模式以及完整模式）的交互分析。



為了控制文章長度與其他文本屬性的共變性，本研究進行迴歸分析時未加入和文本長度相關的文章屬性變項（例如字數、詞數、子句數、詞數等），同時亦將所有可能與文章長度相關的變項常規（normalize）化，例如名詞數、動詞數等詞彙類別屬性，都以文章詞數作為分母把次數屬性轉變為比例變項。除了共變性的考量外，還有兩個原因使得本研究在進行迴歸分析時，控制文章的長度：

（一）本研究採用的目標語料為國小課文，並且以課文所屬年級作為課文的適讀性年級。文章長度與年級的相關是這類語料的特性之一；當然，此特性不必然會出現在國小國語課文以外的語料。其次，在過去適讀性的文獻中，在分析各項與次數相關的文本屬性時，皆會將這些屬性對文章長度做常規化

（normalize）的轉換。例如在 Dale 與 Chall（1948）的公式中，「難詞數」這個屬性是文本中的難詞次數除以文本的總詞數；在 Flesch（1948）的閱讀簡易度模式中，「音節數」亦是計算文本中的總音節數除以總詞數。故本研究亦遵循相似的研究方法，排除與文本長度直接相關的屬性，並將各種次數變項對文章長度作標準化。

（二）在本研究的先期研究中，當與文章長度相關的屬性（如，文本中的字數、詞數、名詞數、名詞片語數、子句數等）納入分析時，年級與這些屬性的相關皆高於.80。這 5 個與文本長度相關的屬性所構成的迴歸模式，已可顯著地解釋文本適讀年級， $F(5, 944) = 551.7, p < .001, R^2 = 0.75$ 。也就是說，文本長度將大幅影響其它與適讀性相關的文本屬性（如字詞頻、一致性屬性、句法複雜度等屬性）在適讀性模式中能貢獻的解釋量。

在上述的考量下，本研究在建構適讀性的迴歸模式時，將控制各種與文章

長度有關的文本屬性。

二、結果與討論



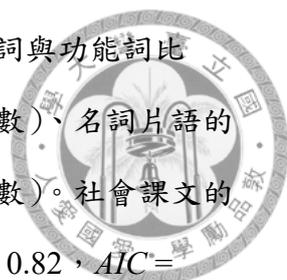
完整迴歸模式結果及屬性係數請參見附表 4.1、4.2 及 4.3。從分析結果中可發現，不管是在全部課文、國語課本或是社會課本中，表層模式的迴歸分析都達到顯著水準。在全部課文中，迴歸分析的結果達到顯著水準， $F(15, 934) = 99.70$ ， $p < .001$ ， $R^2 = 0.62$ ， $AIC = 2813.19$ 。此模式選出 15 個變項，其中 7 個變項的係數達到顯著，分別為子句長（第一四分位數）、句長（第一、第三四分位數）、字頻（第 1、25、50 百分位數）和詞頻（第 1 百分位數）。在國語課本的表層模式中，逐步迴歸選擇出 17 個變項， $F(17, 421) = 43.21$ ， $p < .001$ ， $R^2 = 0.64$ ， $AIC = 1278.00$ 。在這 17 個變項中，共有 7 個變項的係數達到顯著；它們分別為子句長（第 1、3 四分位數）、句長（第 1、3 四分位數）、字頻（第 5、50 百分位數）和詞頻（第 75 百分位數）。在社會課文的表層模式中，逐步迴歸選擇出 11 個變項， $F(11, 499) = 95.71$ ， $p < .001$ ， $R^2 = 0.68$ ， $AIC = 1447.04$ 。在這 11 個變項中，共有 6 個變項的係數達到顯著，分別為子句長（第 1、3 四分位數）、句長（第 3 四分位數）、字頻（第 1、25、50 百分位數）和詞頻（第 25 百分位數）。這個結果顯示，影響國語文本與社會文本之適讀性的文本屬性相同；而這些屬性也與過去的研究發現相似。

Coh-Matrix 模式在全部、國語、社會課文上都比表層模式的解釋力高。全部課文的 Coh-Matrix 模式選擇出 40 個變項， $F(40, 909) = 96.52$ ， $p < .001$ ， $R^2 = 0.75$ ， $AIC = 2440.42$ ，其中 10 變項的係數達到顯著，分別為句長（第 1、第 3 四分位數）、字頻（第 25、50 百分位數）、內容功能詞比例、名詞脈絡重複指標、內容詞的局部與脈絡重複指標、詞彙語意數（第 3 四分位數）和名詞片語的修飾詞數。國語課文的 Coh-Matrix 模式選擇出 38 個變項， $F(38, 400) = 35.75$ ， $p < .001$ ， $R^2 = 0.77$ ， $AIC = 1113.22$ ，其中 13 個變項的係數達到顯著，分別為子句長（第 1 四分位數）、句長（第 1、2、3 四分位數）、詞頻（第 1、

5、75 百分位數)、內容功能詞比例、名詞脈絡重複指標、內容詞的局部與脈絡重複指標、詞彙語意數(第 2 四分位數)和名詞片語的修飾詞數。社會課文的 Coh-Matrix 模式選擇出 31 個變項, $F(31, 479) = 63.67, p < .001, R^2 = 0.81, AIC = 1232.21$, 其中 10 個變項的係數達到顯著, 分別為句長(第 3 四分位數)、字頻(第 25、50 百分位數)、詞頻(第 1、50 百分位數)、主要動詞前詞數、內容詞的局部與脈絡重複指標、詞彙語意數(第 3 四分位數)和名詞片語的修飾詞數。

由於表層模式與 Coh-Matrix 模式經過變數選擇後, 兩者間並非巢套模型(nested models), 故不能直接以 F 檢定比較兩模式, 但從兩模式的 AIC 和解釋力(R^2)仍可發現 Coh-Matrix 模式比表層模式更可解釋文本屬性與適讀年級間的關係。在 Coh-Matrix 適讀性模式中, 國語文本和社會文本的分析結果頗為類似; 除了一些句長、字頻、詞頻等表層文本屬性外, 還納入了影響文本一致性的屬性(如, 內容詞的局部與脈絡重複指標)、句法複雜性屬性(如, 名詞片語的修飾詞數), 以及詞彙複雜度(詞彙語意數)等傳統適讀性公式沒有納入的屬性。兩類文本比較, 國語文本的適讀性受到更多文本一致性的影響, 所以內容功能詞比例、名詞脈絡重複等指標亦達顯著。

完整模式和 Coh-Matrix 模式所納變項頗為相似, 只是前者比後者多納入句法樹深度指標和罕見片語語式頻率等與句法複雜度相關的指標。全部課文的完整模式選擇出 41 個變項, $F(41, 908) = 74.39, p < .001, R^2 = 0.77, AIC = 2374.72$, 其中 13 個變項的係數達到顯著, 分別為子句長(第 2 四分位數)、句長(第 1、3 四分位數)、字頻(第 1、5、50 百分位數)、名詞脈絡重複指標、內容詞的局部與脈絡重複指標、詞彙語意數(第 3 四分位數)、名詞片語的修飾詞數、句法樹最大深度、罕見片語語式頻率(第 3 四分位數)。國語課文的完整模式選擇出 39 個變項, $F(39, 399) = 37.63, p < .001, R^2 = 0.79, AIC = 1088.00$, 其中 13 個變項的係數達到顯著, 分別為子句長(第 1 四分位數)、句



長（第 1、3 四分位數）、詞頻（第 1、5、50 百分位數）、內容詞與功能詞比例、內容詞的局部與脈絡重複指標、詞彙語意數（第 2 四分位數）、名詞片語的修飾詞數、句法樹最大深度、罕見片語語式頻率（第 2 四分位數）。社會課文的完整模式選擇出 39 個變項， $F(39, 471) = 55.03$ ， $p < .001$ ， $R^2 = 0.82$ ， $AIC = 1206.45$ ，其中 14 個變項的係數達到顯著，分別為子句長（第 1、2 四分位數）、句長（第 3 四分位數）、字頻（第 25、50 百分位數）、詞頻（第 1、25、50 百分位數）、主要動詞前詞數、名詞脈絡重複指標、詞彙語意數（第 3 四分位數）、名詞片語的修飾詞數、句法樹最大深度、罕見片語語式頻率（第 3 四分位數）。這些分析結果顯示，無論是哪一類的文本，新增加的兩個與句法複雜度相關的指標均會影響文本的適讀性，被納入適讀性的模式中。

表 2.1 各模式在各文本類型中之迴歸分析結果

文本類型	統計指標	表層模式	Coh-Metrix 模式	完整模式
全部課文 ($N = 950$)	R^2	.62	.75	.77
	AIC	2813.19	2440.42	2374.72
國語課文 ($N = 439$)	R^2	.64	.77	.79
	AIC	1278.00	1113.22	1088.00
社會課文 ($N = 511$)	R^2	.68	.81	.82
	AIC	1447.04	1232.21	1206.45

表 2.1 呈現各文本類型中各模式之迴歸分析結果。從表中可知，完整模式的比表層模式更能解釋文本的適讀性。在各文本類型中，完整模式的解釋力（ R^2 ）皆比表層模式提高約.15，從 AIC 指標亦可看出模式更能描述資料（ AIC 指標數值愈低代表模式較佳）。其次，完整模式和 Coh-Metrix 模式的表現雖然相似，但完整模式的解釋力和 AIC 指標皆比 Coh-Metrix 模式略佳。表 2.2 則進一步整理完整模式中可預測文本適讀性的各個屬性。從表中結果可發現，表層屬性中的子句長、句長、字頻、詞頻在各文本類型中都是能夠預測文本適讀性



的顯著變項。這項結果與過去傳統的適讀性研究一致（胡志偉等人，1994；荊溪昱，1995；陳世敏 1972；Yang, 1970）。在過去研究中亦使用與字頻、詞頻相關的常用字比例、高頻詞數來衡量文章難度，同時亦以句子長度、長子句數等與句長、子句長相關的屬性作為預測公式中的變項。表 2.2 中亦顯現，Coh-Metrix 屬性中之內容詞功能詞比、內容詞重複指標、詞彙語意數、句法複雜度也都在三類文本中可顯著預測文本適讀性。此結果亦呼應過去在中文文本一致性指標的研究。在過去研究中，文本的內容詞數和句法複雜度都是影響適讀性的重要屬性之一（Sung et al., 2015）。本研究則更進一步以句子的句法樹具體操作各項句法複雜度的屬性。從此迴歸分析中，發現與句法樹相關的屬性：名詞修飾詞數、主要動詞前詞數、句法樹最大深度、罕見片語語式頻率都可顯著地預測適讀年級。

研究 1.2 分別針對不同的文本類型計算三種迴歸模式：表層模式、Coh-Metrix 模式以及完整模式。上述模式分析結果顯示，表層模式所涵蓋的變項可以捕捉大部份的文本的實際年級與估計出的文本適讀年級之間的關係。Coh-Metrix 的表現比表層模式略佳。重要的是，Coh-Metrix 屬性可進一步幫助適讀性模式掌握文本句法結構甚至語意的訊息。這樣的結果符合研究者建構適讀性模式時的認知理論考量。最後，本研究的結果除了複製過去傳統適讀性的研究外，亦發現在研究一所計算的各項句法複雜度指標（亦即句法樹深度和罕見片語語式頻率）亦可顯著地幫助模式預測文本的適讀年級。認知理論期研究者也試圖將句法複雜度落實於適讀性模式中，例如，在 Coh-Metrix 中，Graesser 以「主要動詞前詞彙數」間接的推論句法複雜度。本研究所收集的句法樹深度和罕見片語語式頻率等兩個指標更為直接的顯現一個句子的句法複雜度，更重要的是，迴歸分析結果顯示，這兩個指標均被納入適讀性模式中。

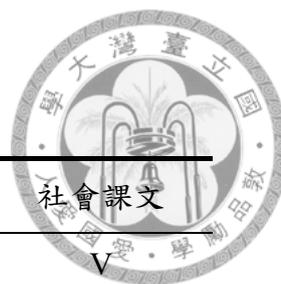


表 2.2 完整模式中顯著預測適讀性之文本屬性

文本屬性	全部課文	國語課文	社會課文
表層屬性	字頻	V	V
	詞頻		V
	子句長	V	V
	句長	V	V
Coh-Matrix 屬性	內容詞功能詞比		V
	名詞重複指標	V	V
	內容詞重複指標	V	V
	詞彙語意數	V	V
	名詞修飾詞數	V	V
	主要動詞前詞數		V
句法	句法樹最大深度	V	V
	罕見片語語式頻率	V	V



第三章 研究二、四種取向建構之適讀性模式

的開發與比較



研究二試圖開發四個不同取向的適讀性模式（線性迴歸模式、支持向量機模式、主題模式和深度學習模式），並比較不同模式對國語及社會文本的預測效能。本研究可幫助研究者評估不同輸入及參數透明度的模式在預測文本適讀性上的表現，以及不同模式預測結果的相關程度。此研究以巢套交叉驗證法（nested cross validation）選擇不同模式的超參數（hyperparameters），並以各模式最佳的超參數來衡量不同模式間的預測正確率。

第一節 研究方法

研究二將以機器學習的方式選擇不同模式的超參數，並衡量不同模式的表現。不同模式各有不同的模式超參數（hyperparameter），為讓不同模式分別選擇較適當的參數以利後續比較，本研究將採用巢套驗證法選擇四模式的參數：分別是線性模式的正規化參數、支持向量機的懲罰參數（penalty parameter）、主題模式的主題數和文本來源、以及深度學習網路的模式架構。

表 3.1 四種適讀性模式

透明度類型	高模式參數透明度	低模式參數透明度
高輸入透明度	線性迴歸	支持向量機
低輸入透明度	主題模式	深度學習

壹、適讀性模式



本研究以輸入透明度與參數透明度區分四種不同適讀性的預測模式（表 3.1）：

一、線性迴歸模式

線性迴歸模式代表高輸入透明度和高參數透明度的適讀性模式。該模式的輸入資料是研究一得到的 45 種文本屬性（77 個指標）。每個文本屬性都已在過去的適讀性文獻中驗證其與文本難度或閱讀歷程的相關，故此模式具備高輸入透明度。另外，此模式的參數即各文本屬性的迴歸係數。雖然多元迴歸模式（相較於簡單迴歸）的係數詮釋較為複雜，但仍在統計上有固定的口語描述方式，故此模式具備高參數透明度。本研究將從研究 1.1 所建立的 77 項文本指標建立一線性迴歸的適讀性預測模式。該模式以 Lasso 法同時進行文本屬性選擇以及參數估計。Lasso 中超參數包含目標函數中的正規化係數。

本研究採用 Lasso 法進行迴歸分析的原因是，研究 1.1 的文本屬性分析顯示，一些文本屬性間存在著高相關。如果本研究以逐步迴歸模式建構適讀性的預測模式，勢必要處理預測變項（regressors）間的多元共線性

（multicollinearity）問題。多元共線性對迴歸分析可能造成的問題包括：（1）整體模式的統計顯著，但變項的係數不顯著，這會影響研究者對係數的詮釋，以及（2）影響逐步迴歸時的變項選擇。事實上，研究一的確可能受到第一問題的部分影響。因為，研究一的各個分析結果均發現多個變項的係數達到統計顯著；這亦符合過去的研究發現。研究一未明顯發現係數標準誤對統計檢定的影響可能來自於迴歸分析的樣本數較大，故共線性問題未嚴重影響到係數檢定。其次，研究一的確發生變項係數較難詮釋的情形：例如，多個分析發現，句長的第 1 和第 3 四分位數的係數方向相反，但研究 1.1 的屬性常模顯現，該屬性應該與文本年級呈現正相關。所幸，此問題可能並未直接影響到迴歸分析結果

(亦即，適讀性模式)的應用價值。

過去研究指出，一種降低多元共線性的方法 (Dormann, et al. 2013) 是直接移除高相關的變項。然而，本研究計算之文本變項不僅提供給傳統適讀性模式的建構，亦提供支持向量機取向的適讀性模式建構。後者是以事例權重計算最後的分類超平面，所以不受變項間的多元共線性的影響。若因為變項間的相關高 (且並非完全相關) 而移除某些文本的屬性變項，會使得以支持向量機取向建構適讀性模式時，減少了輸入特徵的訊息。第二種可能的解決多元共線性問題的方法是先對輸入特徵進行主因子分析 (principal component analysis)，並以主因子當作迴歸分析的預測變項。但是主因子分析後的特徵已非原始的文本屬性，所以這些因子代表的意義已經不具備透明度，同時當分析結果出來後，變項係數的詮釋亦將十分困難；故此方法也不符合本研究的目的。第三種解決多元共線性的方式是使用較耐受共線性的模式，例如 Lasso 迴歸。Lasso 迴歸的優點是可藉由對係數的正規化同時選擇變項和估計變項參數，此方法較符合本研究的目的。

二、支持向量機

支持向量機是高輸入透明度和低參數透明度的適讀性預測模式。大部分採用支持向量機建構適讀性模式的研究都使用文本屬性作為輸入特徵，所以該模式具有較高的輸入透明度。然而，支持向量機所估計的參數為每個訓練模式在空間中做為支持向量的權重，並應用該權重界定出分類平面。此分類過程具有抽象的形式概念，與線性迴歸相比，它的參數透明度較低。本研究以線性支持向量機作為預測模式，並同樣以研究 1.1 的 77 項文本指標建立支持向量機的適讀性預測模式。線性支持向量機的超參數包括目標函數中的懲罰係數 (penalty parameter)。



三、主題模式

主題模式是低輸入透明度和高參數透明度的適讀性預測模式。主題模式計算文章和詞彙的主題分配，並從文本的主題分配預測最可能的適讀年級結果。在主題模式中，由於每個主題都是詞彙的分配構成，故主題本身可由人工事後給予標記（如某些主題是關於環境、或關於自然科學等），亦可提高該演算法在預測上的可訴說性。然而，主題模式的輸入資料，僅是每個文章內的詞彙次數分配，完全與適讀性文獻無關，故在輸入資料透明度上是較低的。

在採用主題模式取向建構適讀性模式的過程中，會根據訓練文本建立多個主題，也會建立每篇文本的主題表徵。不僅如此，同時該模式也可計算一篇（不在訓練文本中的）新文本的主題表徵。本研究將從語料中建立主題模式（Rehurek & Sojka, 2010），並比較每篇文章的主題表徵，找出所有訓練文本中與新文本主題最接近的文本表徵當作新文本之適讀年級的預測。亦即，最後預測結果是以 k-近鄰演算法（k nearest neighbor, kNN）作為決定預測結果的方式。

主題模式的超參數選擇包括（1）訓練資料來源、（2）主題數。主題機率模式本身的事先機率（priors）參數，亦即文章主題分配和主題詞彙分配的狄氏分配（Dirichlet distribution）機率參數，則是由語料中自動推估出不對稱的狄氏分配參數（Wallach, Mimno, & McCallum, 2009）。本研究在交叉驗證過程中，主要評估不同語料來源和主題數對適讀性預測的表現。本研究的語料來源考慮平衡語料庫以及小學國語、社會課文等兩種語料類型。在不同語料類型下，本研究分別比較主題數為 50、100、200、300、400 的預測正確率。

四、深度學習

深度學習模式是低輸入透明度和低參數透明度的適讀性預測模式。深度學習模式的模式參數包含每個神經節點的連結權重，這些參數可理解為神經網絡

用來逼近任意函數的工具 (Liang & Srikant, 2017)，但在實際應用的問題層面較難找到可訴說的方法理解參數意義。在模式的輸入特徵方面，深度學習在語言分類作業常以詞彙作為網路輸入，故深度學習的適讀性模式可不需要任何文本屬性，其輸入透明度上是較低的。



深度學習在文本分類作業上有相當多選擇，本研究考量到所有的文本資料僅有 950 篇，甚至還需要區分國語與社會課文，故本研究僅使用較單純的神經網絡。

本研究使用的深度學習模式之輸入資料為一篇篇課文斷詞後的原始詞彙。所有課文中的詞彙都先透過一個事先定義的詞典 (dictionary)，該詞典將每個詞彙對應到一個正整數。故每篇文章的詞彙都可被表達為一組長度為 N_d 的輸入向量 (N_d 為該篇文章的詞數)。這組輸入向量會接著進入「嵌入層」(embedding layer)。嵌入層是包含 $V \times N_{dim}$ 矩陣的查找 (lookup) 網路層。 V 代表輸入向量的詞數範圍，本研究使用中研院平衡語料庫為詞典來源，共有 28,272 詞； N_{dim} 是嵌入層的維度數，亦即，一個詞彙需用到幾個數字表達。嵌入層會一個一個把輸入向量的正整數當作索引數值，從嵌入矩陣中找出對應該詞彙的向量。經過嵌入層後，每個詞彙將會由純量轉變為的向量 (向量程度為 N_{dim})，亦即 N_d 的輸入向量經過嵌入層後，會變成一個二維矩陣 ($N_d \times N_{dim}$)。

接著此二維矩陣需經過進一步編碼 (encode) 成一個文章向量，模式可藉助此文章向量做各類型的文本歸類作業 (如此研究的適讀性預測作業)。在目前深度學習的文獻中，此步驟有諸多可能的選擇：平均化層 (average pooling)、循環層 (recurrent layer)、長短期記憶層 (long short-term memory, LSTM)、摺積層 (convolutional layer) 等。然而，循環層當序列長度較長時，很容易會有梯度消失的問題；此研究所使用的訓練材料皆來自於一篇文章，文章長度在中高年級可能到達上千詞，使用 RNN 在訓練上並不容易。而長短期記憶層與摺積層所需要訓練的參數數目眾多，此研究的材料僅 950 篇，且仍須區份國語科

與社會科文本，訓練材料尚不足以滿足參數訓練所需。故本研究使用平均化層，將 $N_d \times N_{dim}$ 的矩陣，沿著矩陣的行平均起來成一組向量（長度為 N_{dim} ），當作此模式接續預測的文本向量。

當取得文本向量，本研究直接將此文本向量藉由前餽層（feed-forward layer）連結至輸出層（output layer）。此前餽層是完全連接（fully connected），並以 sigmoid 函數作為激發函數（activation function）。文本向量中的每個元素，都可直接影響每個輸出節點。輸出層共有 6 個節點，分別代表 1 至 6 年級。輸出節點的值將經過 softmax 轉換，讓適讀性的預測符合機率分配的特質。最後再將預測結果和實際文本在課文中的年級做比較。由於每個文本僅屬於一個類別，而模式的目的是使正確的類別有最大的機率，故本研究選用的目標函數為負對數似然函數（negative log-likelihood function）。

整體而言，本研究採用的深度學習模式架構為詞彙輸入連結至詞彙嵌入層，並使用平均化層降低維度（average pooling），並連結至輸出層形成預測。在此架構中，嵌入層的嵌入維度數目（20、50、100、200）是待決定之模型超參數。

此四模式所預測的文章適讀性皆未使用到與文章長度相關的訊息。高輸入透明度之線性迴歸模式與支持向量機模式皆使用研究一所計算之 77 個文本屬性，這些屬性皆已去除與文章長度相關變項或對變項做標準化，故此二模式皆已在輸入資料的階段控制文章長度訊息。低輸入透明度的主題模式的輸入資料是每篇文章的詞彙機率分配，故每篇文章的長度已在該機率分配中被標準化（normalized），此步驟相當於在輸入層次控制了文章長度。在深度學習模式中，雖然文章長度與輸入層的維度相關（ $N_d \times N_{dim}$ ），但在本研究的深度學習架構中，輸入層的訊息經過嵌入層編碼後，就直接進入平均化層（average pooling）降低維度成一組長度為 N_{dim} 的文章向量表徵。亦即文章長度亦已在平均化層被控制了。故此研究使用的四個模式皆已控制文章長度的因素。

貳、模式參數選擇



本研究使用巢套交叉驗證法 (nested cross validation) (Stone, 1974; Cawley & Talbot, 2010) 進行參數選擇與模式評估。此方法的優點在於模式參數選擇與結果評估使用不同的資料，可降低在模式因參數選擇與評估的資料重複而產生的正確率偏差 d 。

巢套該驗證法分成外層與內層交叉驗證，前者之目的在於評估模式結果，後者之目的為選擇模式之超參數。本研究將資料分成 90% 的驗證資料，作為內層交叉驗證使用；以及 10% 的測試資料，作為外層驗證中最後的模式評估的資料。在內層交叉驗證中，驗證資料將按各年級比例分成 5 份，並做 5 次的模式驗證。在每次驗證中，驗證資料中的 1 份將輪流作為測試資料，其他 4 份則作為訓練資料，每次驗證都將得到 1 個驗證的訓練分數，和驗證的測試分數。當內層驗證完成後，每個模式都可以計算內層驗證的訓練正確率和測試正確率的平均數和標準差。

本研究按照內層交叉驗證結果選擇表現最佳的參數，進入外層的交互驗證。在外層交互驗證程序中，模式以 90% 的測試資料（亦即所有的內層驗證資料）作為訓練資料，並以剩下 10% 的資料作為測試資料。這組測試資料是模式在內層交互驗證做參數選擇時完全沒有「看」過的，該正確率即為模式最後的評估結果。

本研究共分成三種語料類型，分別是國語課文、社會課文和全部（國語和社會）課文。此三類語料分別都經過上述巢套交叉驗證的程序。在每個語料類型中，內層交互驗證的參數選擇階段中，需要選擇的模式參數共有 20 組，分別是：線性迴歸模式的 3 組參數（正規化係數）、支持向量機的 3 組參數（懲罰係數）、主題模式的 10 組參數（主題語料來源與主題數）和深度學習的嵌入層維度（4 組）。

進入外層交叉驗證的模式評估階段後，上述各模式都將選擇出一組最佳參數。故各語料類型（國語、社會、全部課文）各有 4 組模式比較模式表現。



第二節 結果與討論

研究二透過巢套交叉驗證選擇出四組模式最佳的超參數，並比較四組模式在三種不同類型文本（國語、社會、全部課文）的正確率。

壹、參數選擇結果

研究二共比較四種不同模式，共有 20 組不同的模式參數。每組模式參數都在三種不同類型的語料中，經過 5 次不同的交叉驗證得到 5 個訓練正確率、以及 5 個測試正確率。所有交叉驗證的結果整理如表 2.4。以下將分別討論各種不同模式在不同參數下的驗證正確率結果。

在線性迴歸模式中，在不同的文本類型中，均呈現正規化係數愈低，測試正確率愈高的趨勢。在全部課文的語料類型中，當正規化係數為 0.05 時，測試正確率為.44；係數為 0.10 時，正確率為 0.42，當係數為 0.20 時，正確率為 0.36。在國語課文的語料類型中，當正規化係數為 0.05 時，測試正確率為.43；係數為 0.10 時，正確率為 0.38，當係數為 0.20 時，正確率為 0.31。社會課文的語料類型中，當正規化係數為 0.05 時，測試正確率為.47；係數為 0.10 時，正確率為 0.43，當係數為 0.20 時，正確率為 0.39。本研究最後選擇在所有文本中具有最高測試正確率的正規化係數（0.05）作為線性迴歸的超參數。。

在支持向量機模式中，三種模式的超參數亦有接近的正確率表現。支持向量機所比較的參數為懲罰係數為 0.01、0.1、和 1.0 時之模式正確率表現。在全部課文的語料類型中，當懲罰係數為 0.01 時，測試正確率為.50；係數為 0.1 時，正確率為.38；係數為 1.0 時，正確率為.48。在國語課文中，在各懲罰係數

下之正確率皆為.38。在社會課文中，當懲罰係數為 0.01 時，測試正確率為.61；係數為 0.1 時，正確率為.61；係數為 1.0 時，正確率為.59。本研究最後選擇之懲罰係數 0.1 作為最佳的模式超參數。



在主題模式中，共有 10 組不同的超參數組合，分別是 2 種不同的主題模式訓練語料來源；以及不同語料訓練來源中，分別有 5 種不同的主題數。從交叉驗證結果中可發現，不管是在哪一種文本類型（國語、社會、全部），以課文當作主題模式之訓練來源的測試正確率，均低於以平衡語料庫作為訓練語料來源之正確率。在平衡語料庫的主題模式中，主題數愈多的預測表現會愈好，故最後本研究選擇以平衡語料庫建立主題數為 400 的主題模式作為最佳模式。

上述結果透露出主題模式的來源語料與適讀性預測的有趣關係。主題模式的訓練語料某種程度上代表其主題分配所在的語義空間，就直觀上而言，此預測作業既然是預測課文的適讀性，以課文本身當作主題模式的來源應有助於預測作業，但交互驗證的結果並不符合這項預期。然而，主題模式的估計過程需從詞彙的出現頻率估計主題的詞彙分配，故詞彙本身在語料庫的數量也和主題空間的穩定性息息相關。以語料庫大小而言，平衡語料庫是百萬字規模的語料庫，但全部課文僅約 30 萬字。兩者的語料規模可能對主題模式應用在適讀性預測作業的結果有很大的影響。

深度學習模式有 4 項超參數，亦即詞彙嵌入層需要多少嵌入維度。在四組超參數（維度為 20、50、100、200）中，維度為 100 時的測試正確率表現高於其他參數。維度 100 之參數在全部課文的測試正確率為 0.57、在國語課文中為 0.41、在社會課文中正確率為 0.75。故最後深度學習模式以維度 100 作為最佳之嵌入層維度。

貳、模式正確率



經過內層交叉驗證的參數選擇階段後（交叉驗證結果請見表 3.2），本研究共選出 4 組最佳超參數分別代表上述四個模式：分別是線性迴歸的正規化係數使用 0.05；支持向量機的懲罰係數為 0.1；主題模式使用平衡語料庫為訓練語料，主題數設定為 400；深度學習之嵌入層維度為 100。

此 4 組模式進入外層的交叉驗證模式評估階段。在這階段中，各模式使用內層交叉驗證的全部資料作為訓練資料，並預測從未「看過」的測試資料。各模式在測試資料的正確率整理如表 3.3。

此四模式都分別計算它們在不同文本類型（國語、社會、全部課文）的訓練和測試正確率。在全部課文的文本類型中，正確率最高的是深度學習（ $M = 0.62$ ）、其次是支持向量機（ $M = .56$ ），主題模式的正確率則較低（ $M = 0.37$ ）。線性迴歸的正確率亦較低（ $M = 0.38$ ），該模式選擇了 32 個變項，包括字頻（第 1、50 百分位數）、字頻排序（字頻排序在 1-800、3000 以上的字數比例）、字筆畫數（第 2、3 四分位數）、子句長的中數、內容詞重複指標（局部與脈絡）、名詞局部重複指標、句長（第 3 四分位數）、詞長（第 2 四分位數）、詞頻排序比例（在 5 萬到 10 萬間）、句法樹最大深度、罕見片語語式頻率（第 3 四分位數）、名詞修飾詞數、詞彙語意數（第 2、3 四分位數）、形容詞比例、被字句比例、連接詞（因果、目的、轉折、遞進）比例、動詞（所有動詞、能願動詞、使令）比例、名詞比例、代名詞比例、詞類詞例比、動詞片語比例、和動詞比例。

在國語課文的文本類型中，正確率最高的是線性迴歸（ $M = 0.55$ ），其次是深度學習（ $M = 0.48$ ），次之是支持向量機（ $M = 0.41$ ）和主題模式（ $M = 0.32$ ）。線性迴歸模式在國語課文中選擇出 27 個變項，分別是字頻排序（排序在 800 以下、和排序在 3000 以上）、字筆畫數（第 1 四分位數）、子句長（第 1

四分位數)、內容重複指標(局部與脈絡)、名詞重複指標(局部)、句長(第1、3四分位數)、詞頻(第75百分位數)、詞長(第2四分位數)、詞頻排序比例(排序在10萬以上)、最大句法樹深度、名詞修飾詞數、詞彙語意數(第2、3四分位數)、連接詞(全部、因果、解證、目的、假設、轉折、選擇)比例、內容詞功能詞比、名詞片語比例、詞類詞例比和動詞片語比例。

在社會課文的文本類型中,正確率最高的是深度學習($M=0.77$)、其次是同屬低輸入透明度的主題模式($M=0.67$)。高輸入透明度的支持向量機的正確率次之($M=0.54$)、線性迴歸的正確率則較低($M=0.48$)。線性迴歸模式在社會課文中選擇了28個變項,包括字頻排序(字頻排序在1-800)、內容詞重複指標(局部、脈絡)、名詞局部重複指標、語意重複指標(局部)、句長(第3四分位數)、詞長(第2四分位數)、詞頻排序比例(在5萬到10萬間)、句法樹最大深度、罕見片語語式頻率(第3四分位數)、名詞修飾詞數、詞彙語意數(第2、3四分位數)、主要動詞前詞數、形容詞比例、被字句比例、連接詞(因果、目的、轉折、遞進、選擇)比例、內容詞比例、動詞(能願動詞)比例、名詞比例、代名詞比例、詞類詞例比、動詞片語比例、和動詞比例。

從上述結果中,深度學習的結果相較於其他模式而言,在各種文本類型的表現較為穩定。而線性迴歸模式在國語課文有最高的正確率,但在全部課文的預測正確率卻較差。同屬高輸入透明度的支持向量機模式在三種不同文本類型也較穩定,但在國語課文的正確率最低。而主題模式在國語課文和全部課文的表現都較差,但在社會課文卻有僅次於深度學習模式的正確率。

模式歸類的正確率提供很容易判讀的訊息,但模式預測錯誤的結果也暗示關於模式的預測特性。尤其適讀性年級的預測雖然是文本歸類問題,但不同適讀年級並不是完全獨立的類別變項。不同適讀年級的預測應該具有一定程度的順序關係。在正確率指標的計算上,模式將一篇5年級的文章預測成6年級,和將該文章預測成1年級,在正確率指標上都同樣是「錯誤」的,但兩者在適

讀性應用上卻有截然不同的意義。圖 3.1 即試圖以各模式在不同文本類型的混淆矩陣 (confusion matrix) 呈現各模式的預測適讀性和真實文本年級之間的相關。



圖 3.1 混淆矩陣的橫軸是真實的文本年級，縱軸是各模式 (在不同文本類型下) 所預測的適讀年級。如果預測結果是完全正確的，則混淆矩陣的對角線應該是黑色的方塊，例如圖 3.1 右下角的混淆矩陣代表深度學習模式預測社會課本的結果，是最接近完全正確的結果。從混淆矩陣的結果可以發現，雖然線性迴歸模式在社會課文的預測表現不高，但從混淆矩陣的結果而言，模式最多錯誤的地方是無法分辨低年級的文本，而對三到六年級的文本的預測，則可能會上下相差 1 個年級。此錯誤樣態符合社會課文是傳遞領域知識的學科，除了低年級生活和中高年級社會科行文風格確有不同外，在社會科不同年級文本的主要差異較傾向顯現於不同領域的主題 (如地理、歷史、經濟、政治) 等。此特性亦顯現在主題模式較能掌握社會科適讀年級的結果。從混淆矩陣也可看出主題模式對社會科的預測正確率較高，僅在二年級的生活科文本犯了較多錯誤，這和低年級文本屬於生活領域，在文本主題上較不明確有關。

從混淆矩陣的結果中，線性迴歸模式在國語課本中的確有較佳的表現。該模式大部份的錯誤是把許多一年級文章預測成二年級，以及混淆許多高年級的文本，暗示高年級國語課文的行文難度已較為接近。過去研究 (曾昱翔等人，2014) 亦發現高年級國語課文的用字型態已逐漸接近。另外，從混淆矩陣亦可發現，雖然深度學習在國語課文上的預測正確率較差，但其預測錯誤大部份僅與實際文本年級相差一到二年級。由於深度學習模式的輸入特徵是詞向量表徵，詞彙嵌入層較能掌握的是詞彙共現關係 (與詞彙語義較有關係)。不同年級國語課文的難度更傾向發生在用字用詞、句法結構等屬性上。此研究用的深度學習模式未納入對語句結構較敏感的網路架構，故深度學習模式應難以掌握到和句法相關的屬性。但目前深度學習發展相當成熟，只要資料量足夠，深度學

習預測國語課文的正確率仍有提升的可能。

從文本真實年級與模式預測年級的相關係數亦可觀察到與混淆矩陣一致的訊息。該相關結果與各模式的相關矩陣整理如圖 3.2。圖中呈現的相關係數為各模式預測彼此之間的相關，以及和文本實際所在年級中的相關。從該圖之結果可見，雖然各模式的正確率可能不同，但它們和真實的文本年級都有很高的相關。例如國語課文的文本年級與主題模式預測的相關是.75、與深度學習預測的相關是.81；社會課文的文本年級與線性迴歸的相關是.86、與支持向量機預測的相關是.74。模式彼此之間的預測亦有顯著相關，尤其同屬高輸入透明度的線性迴歸與支持向量機的相關在三個不同文本類型中皆在.75 以上；同屬低輸入透明度的主題模式與深度學習的相關亦在.60 以上。高輸入、高參數透明度的線性迴歸模式和低輸入、低參數透明度的深度學習模式之間，在國語課文， $r = .78$ 、社會課文， $r = .87$ 、全部課文， $r = .73$ （以上所提的關係數均達 $p < .001$ 的顯著相關）。

整體而言，研究二開發的四種適讀性模式，深度學習整體表現較佳，在全部課文和社會課文裡都有最高的預測正確率。線性迴歸模式雖然表現稍微較低，但在國語課文中更有比深度學習更高的正確率。以文本主題內容來預測適讀年級的主題模式，在預測社會課文類型的文本時，正確率表現僅次於深度學習，但在預測國語課文時表現較為不佳。此差異的背後原因，可能來自於主題模式預測適讀年級的方式仰賴於不同年級文本需要明顯和不同主題有關。故當不同年級文本有明顯的主題結構差異時，主題模式較能捕捉不同年級的內容區別。相對地，國語課文的差異多在用字遣詞和語句複雜度上，所以這個模式對國文文本適讀性的預測表現不佳。



本研究亦發現不同輸入或參數透明度的模式預測國小課文適讀年級的正確率雖然不同，但彼此都有相當高的相關。此結果顯示，在目前研究使用的課文（國小一到六年級的國語與社會科）中，不同的適讀性預測模式都足以在應用上發揮功能。但是不同模式透明度的差異，是否會影響使用者在應用模式結果時的信任程度，則是研究三試圖探討的問題。



表 3.2 四種模式之交叉驗證結果

模式參數	全部文本		國語文本		社會文本	
	訓練	測試	訓練	測試	訓練	測試
線性迴歸						
alpha005	.46(.01)	.44(.03)	.48(.01)	.43(.02)	.50(.01)	.47(.06)
alpha010	.42(.01)	.42(.03)	.42(.02)	.38(.04)	.45(.01)	.43(.08)
alpha020	.36(.02)	.36(.06)	.33(.01)	.31(.04)	.41(.01)	.39(.05)
支持向量機						
C001	.65(.01)	.50(.02)	.70(.02)	.38(.02)	.81(.01)	.61(.02)
C010	.70(.01)	.49(.02)	.77(.02)	.38(.02)	.90(.01)	.61(.02)
C100	.72(.01)	.48(.02)	.81(.03)	.38(.03)	.94(.01)	.59(.03)
主題模式						
TB_T050	.59(.01)	.31(.02)	.59(.01)	.28(.04)	.63(.01)	.31(.07)
TB_T100	.59(.01)	.27(.05)	.58(.01)	.23(.05)	.66(.01)	.31(.05)
TB_T200	.61(.01)	.25(.05)	.55(.02)	.20(.05)	.62(.01)	.30(.04)
TB_T300	.58(.01)	.21(.05)	.52(.02)	.17(.01)	.57(.01)	.28(.06)
TB_T400	.60(.01)	.26(.03)	.48(.02)	.21(.05)	.60(.01)	.30(.02)
Asbc_T050	.65(.02)	.38(.04)	.59(.03)	.24(.05)	.74(.01)	.51(.04)
Asbc_T100	.66(.02)	.37(.03)	.59(.02)	.27(.04)	.74(.01)	.49(.05)
Asbc_T200	.66(.01)	.38(.04)	.62(.02)	.26(.03)	.72(.01)	.47(.09)
Asbc_T300	.66(.01)	.38(.04)	.62(.02)	.26(.03)	.72(.01)	.47(.09)
Asbc_T400	.71(.01)	.43(.04)	.67(.02)	.33(.04)	.76(.01)	.54(.08)
深度學習						
NN_E200	.84(.01)	.54(.04)	.89(.12)	.37(.08)	.79(.04)	.70(.11)
NN_E100	1.00(.00)	.57(.04)	1.00(.00)	.41(.06)	.96(.09)	.75(.03)
NN_E050	.96(.07)	.55(.02)	1.00(.00)	.41(.08)	.85(.08)	.71(.07)
NN_E020	.86(.08)	.56(.05)	1.00(.00)	.39(.08)	.81(.01)	.71(.06)

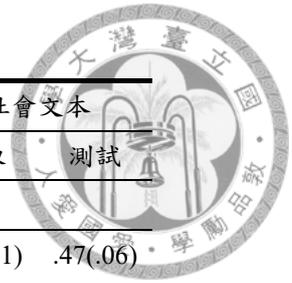


表 3.3 各適讀性模式正確率

高輸入透明度	高參數透明度		低參數透明度	
	線性迴歸		支持向量機	
	訓練	測試	訓練	測試
國語課文	.47	.55	.74	.41
社會課文	.48	.48	.87	.54
全部課文	.46	.38	.69	.56
低輸入透明度	主題模式		深度學習	
	訓練	測試	訓練	測試
	國語課文	.67	.32	.99
社會課文	.76	.67	.82	.77
全部課文	.70	.37	.99	.62

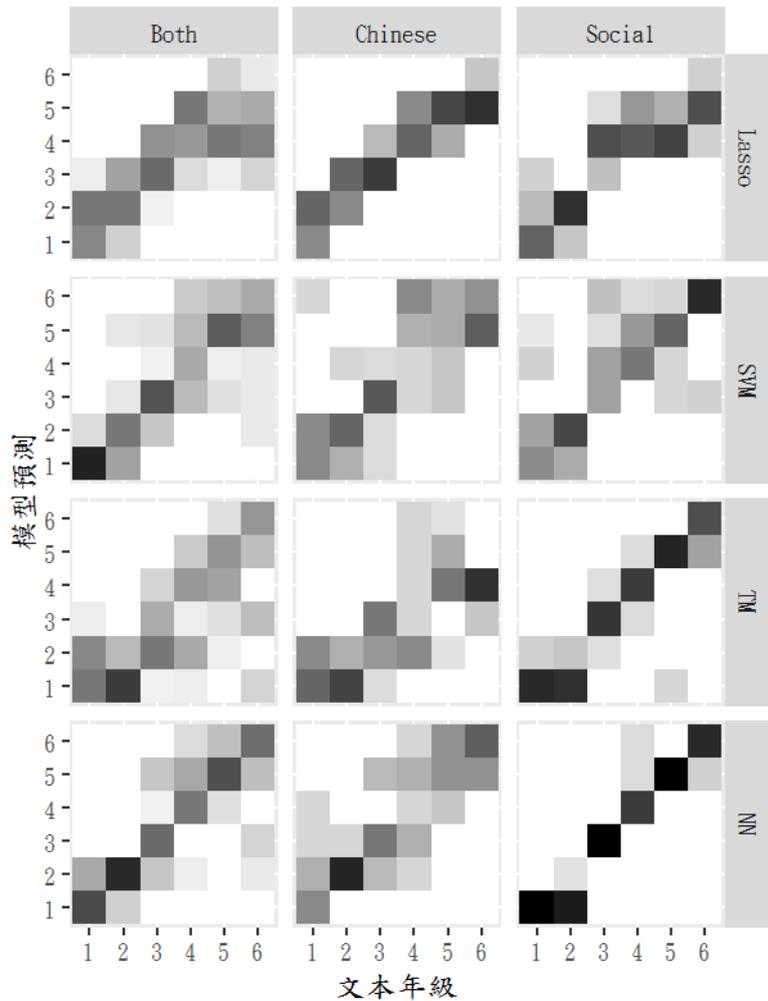


圖 3.1：各模式在不同文本類型下預測之混淆矩陣

全部課文					國語課文					社會課文				
	L	S	T	N		L	S	T	N		L	S	T	N
Lasso	1				Lasso	1				Lasso	1			
SVM	.82	1			SVM	.75	1			SVM	.84	1		
TM	.63	.59	1		TM	.70	.55	1		TM	.84	.73	1	
NN	.73	.79	.60	1	NN	.78	.53	.65	1	NN	.87	.79	.92	1
Grade	.79	.82	.71	.83	Grade	.92	.72	.75	.81	Grade	.86	.74	.91	.96

圖 3.2：各模式預測在不同文本類型下之相關矩陣



第四章 研究三、四種不同模式信任度比較



研究三之目的是探討小學現場教師對於四種不同適讀性模式的看法。此研究以一對一訪談並輔以紙本問卷的方法，讓小學教師瞭解預測適讀性的四種不同建構取向，並詢問他們對於不同適讀性模式的信任程度。

第一節 研究方法

壹、參與者

此研究的參與者是 31 位小學教師。其中 10 位任教於北部學校、21 位來自於中部學校。教學的平均年資為 18.68 年，標準差為 6.43 年。

貳、問卷與施測程序

研究者與參與者是一對一於（參與者任教學校中的）獨立空間中進行研究訪談。參與者在研究開始時會得到一份 A4 列印的紙本問卷，該問卷包含指導語、研究參與同意書、以及問卷內容。研究者會向參與者表明這是一項學位論文研究，主題是「國小文本適讀年級」，研究目的是「想要知道教學現場的教師，認為哪些預測文本適讀年級的方式是他們願意信任的」。研究者徵得參與者同意並簽署同意書之後，開始進行訪談。在徵得參與者同意之後，研究者將訪談過程加以錄音記錄。

訪談時，研究者會先向教師介紹目前預測適讀性的四種建構取向。研究者先以口語方式，敘述四種模式如何預測文本的適讀年級：

「第一種方式和第二種方式根據的都是文章的屬性，像是字、詞、句子有



多難、文章有多長。一和二不同的地方，在於一是比較簡單的數學式子，我們會調整每個屬性的加權，並得到最後的結果。二則是比較複雜的數學模式，每篇文章都會對應到某個區間、這些區間再對應到某個年級。三則是看主題，例如有篇文章和勞動與勞工有關，那就看這個主題和哪個年級的課文主題比較接近。四則是比較接近現在 AI 或是深度學習的作法，我們交給電腦一些文章，以及這些文章的所屬年級，讓電腦自己學習文章和年級的對應關係。」

研究者亦會指明，在紙本問卷中，有各種適讀性模式的文字說明。以下是問卷中對於四種模式的描述：

第一種模式是根據文本所用字詞的難度、句子的長度、文法的複雜性等文本屬性，以統計迴歸分析的方式，評估文本的適讀年級。

第二種模式是根據文本的屬性，建立以屬性為維度的多維空間。然後建立數學模型，將空間區分為不同的區間，每一區間代表不同的適讀年級。

第三種模式是分析文本中每個詞彙的意義，然後藉以推斷文本涉及的內容（主題、領域等）以及各類內容在文本中所佔的比重，藉以評估文本的適讀年級。

第四種方式是提供電腦多篇不同年級的文本，讓電腦建立文章中的詞彙與文本適讀年級的數學關係模型，藉此模型評估其它文本的適讀年級。

當研究者說明完四種不同的模式建構取徑之後，研究者接著請參與者按照他們對這四種模式的大略理解，教師在選擇和評估文章的難度時，「對上述四種模式，你最願意採用哪一種模式的評估結果？請根據您願意採用的程度，對四種模型加以排序。」在問卷中以及研究者的口語敘述中皆未提及四種模式的正確率數值或其相對表現。

參與者需在問卷上，以排序的方式表達他們對於這四種模式採用的願意程度。當參與者作答完畢之後，研究者會以口語詢問參與者排序的理由，參與者則以口語回答。

第二節 結果與討論



問卷結果如圖 4.1，參與者對四種模式的信任度排序依序為主題模式 ($M = 1.84$, $SD = 1.00$)，支持向量機 ($M = 2.58$, $SD = 0.84$)，線性迴歸 ($M = 2.71$, $SD = 1.18$)，以及深度學習 ($M = 2.87$, $SD = 1.18$)。這四個模式的排序分數經由 Friedman 排序總和檢定 (Friedman rank sum test) 達到統計上的顯著差異 ($\chi^2(3) = 11.63$, $p < .01$)。

從參與者口頭報告的內容可歸納出以下的排序原因：(1) 選擇主題模式的教師多認為模式應考慮到文章的內容 (主題)，因為文章內容才是讓學生能夠有興趣閱讀的因素，也才能避免學生學到零散的知識。否則，學生很可能認得文章中的每個字詞，但無法理解文章背後要傳遞的題材。(2) 選擇線性迴歸或支持向量機的教師，傾向認為根據字、詞等文本屬性預測文本適讀性比較接近教師平常以人工選擇文本的方式。而且他們認為，句子的長度、句法複雜度與學生的閱讀理解能力息息相關，較符合教學上的需求。(3) 選擇深度學習的教師則認為電腦在訓練資料中可能發現人們未思考到的面向。而且深度學習會從較多資料中學習年級與文本的關係，相較而言，線性迴歸與支持向量機的想法則過於簡單。

綜合研究二和研究三的結果可發現，各適讀性模式的預測準確率皆具備應用上的價值，但是參與者對於模式的採用程度卻會受到不同模式透明度的影響。研究三問卷訪談的結果顯示，會考慮文本的內容、文章字詞難度、句法複雜度的適讀性模式是參與者較願意採用的模式，而這些特性不見得是所有建構適讀性模式的取向都具備的。

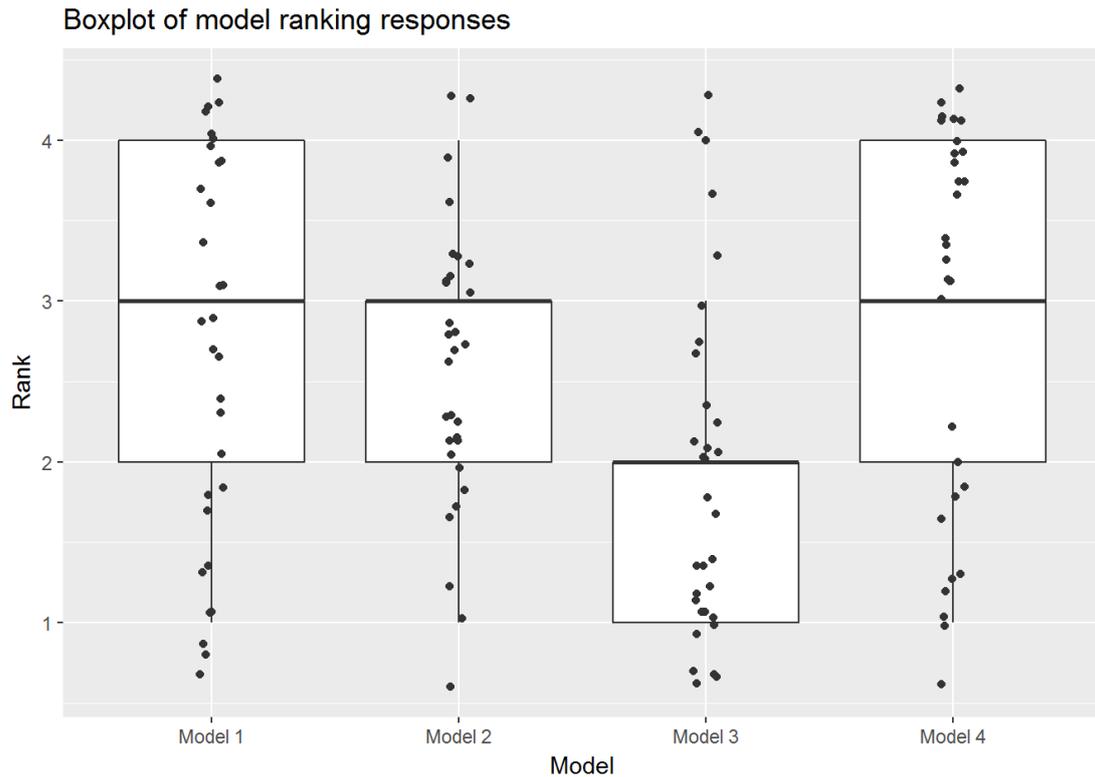


圖 4.1：小學教師對四種不同模式願採用程度之排序

註：數字愈小代表愈信任。模式一到四依序為線性迴歸、支持向量機、主題模式以及深度學習。為清楚呈現各資料點的散佈趨勢，故資料點位置以打散方式 (jitter) 呈現。

第五章 研究四、建立適讀性的診斷系統



研究四之目的是建立一套文本適讀性的診斷系統。當使用者以這系統評估一篇文本的適讀性時，這套系統將整合不同模式對這篇文本之適讀性的分析與預測，然後參考研究三的小學教師對適讀性模式的看法，以即時視覺化的方式，選擇性的將系統對一篇文本分析的結果呈現給使用者參考。研究者通過這樣的設計，能讓使用者能更願意採用本系統對文本適讀性的預測結果。

從研究二的結果中發現，在本研究開發的四種適讀性模式中，正確率最高的是深度學習模式；但是從研究三的結果顯示，該適讀性模式的預測結果卻是小學教師較不願意採用的過去研究認為，使用者對系統的信任程度會受到系統透明度的影響，尤其當預測結果不符合使用者預期，使用者會更需要其他訊息才願意接受系統所提供的資訊 (Kizilcec, 2016)。研究三的結果顯現，不同模式對於適讀性的預測雖不完全相同，但彼此間皆有顯著高相關。所以，診斷式系統可藉此特性，跨越各模式的界線，整合各模式的輸入資料及參數，以較全面的角度提供使用者各層次的文本適讀性診斷訊息。

本系統選擇三個適讀性模式：深度學習模式、線性迴歸模式和主題模式，作為診斷式系統預測文本適讀年級的模式。深度學習模式是在研究二中表現最佳的模式，它在所有文本類型中有最高的正確率。線性迴歸雖然表現不若深度學習，但在研究三中，教師仍認為文本屬性是人工選擇文章時會考量的因素，而線性迴歸模式是與文本屬性最直接相關的適讀性模式。參與研究三的教師認為文本的主題與文本的適讀性最有關係，所以主題模式分析出的結果也將在本系統中扮演重要的角色。

當使用者使用本系統分析一篇文本的適讀性時，本系統將會呈現深度學習模式對文本適讀性的預測值，也將會呈現線性迴歸模式對文本之屬性的分析。



對於使用者對模式之預測結果的信任的問題：(1) 當適讀性模式的預測結果符合使用者的預期時，即便深度學習模式的預測過程未使用與文本內容、字詞難度、句法複雜度等文本屬性，但本系統呈現的文本屬性資料，不會影響使用者對文本適讀性預測的信任。(2) 當適讀性模式的預測結果未符合使用者預期時，本系統可藉由提供各種文本的屬性資料，以及文本的主題相關資料，讓使用者瞭解該這篇文本的屬性為何，以及主題為何。如此，或可提升使用者信任系統對文本適讀性的預測，因為他們知道了，是哪些文本屬性影響了文本的適讀性。更有甚者，使用者可以根據系統提供的訊息，當場在電腦上修改文本，並將新文本輸入系統，檢視修該後的文本的適讀性為何。

本研究建立之系統是可直接以瀏覽器連結的網頁程式，然後透過與後端伺服器的自然語言演算法和適讀性預測模式的整合，使得本系統能夠提供以下功能：(1) 預測文本適讀年級；(2) 分析文本屬性；(3) 協助使用者連結文本適讀性與文本字、詞、句子之間的關係；(4) 提供與該文本內容相似的其他文章。

第一節 系統架構

此系統的整體架構可大致分為六個處理步驟（如圖 5.1）。使用者首先 (1) 在網頁介面輸入欲分析的文本之後，文本內容會送至運算程式。該運算程式 (2) 應用 Stanford CoreNLP (Manning et al., 2014) 的自然語言處理管線完成斷詞、詞類標記、句法剖析三項分析；接著 (3) 藉由研究一的分析結果抽取出 77 項文本屬性。透過上述步驟所得到的文本屬性和斷詞後的詞彙訊息，(4) 進入研究二開發的預測模式，得到 3 種和適讀性預測相關的訊息：(i) 從線性迴歸模式，計算出「文本屬性模式」的預測適讀年級；(ii) 從深度學習模式，得到「深度學習模式」的預測適讀年級，以及 (iii) 從主題模式估計文章的主題

分配，並藉此在系統內建的文章（本研究使用的社會科文本）中尋找最相似的文本。

運算程式完成上述計算作業後，系統接著會將結果（5）傳回網頁介面，由網頁介面（6）將結果以文字及視覺化的方式呈現給使用者。

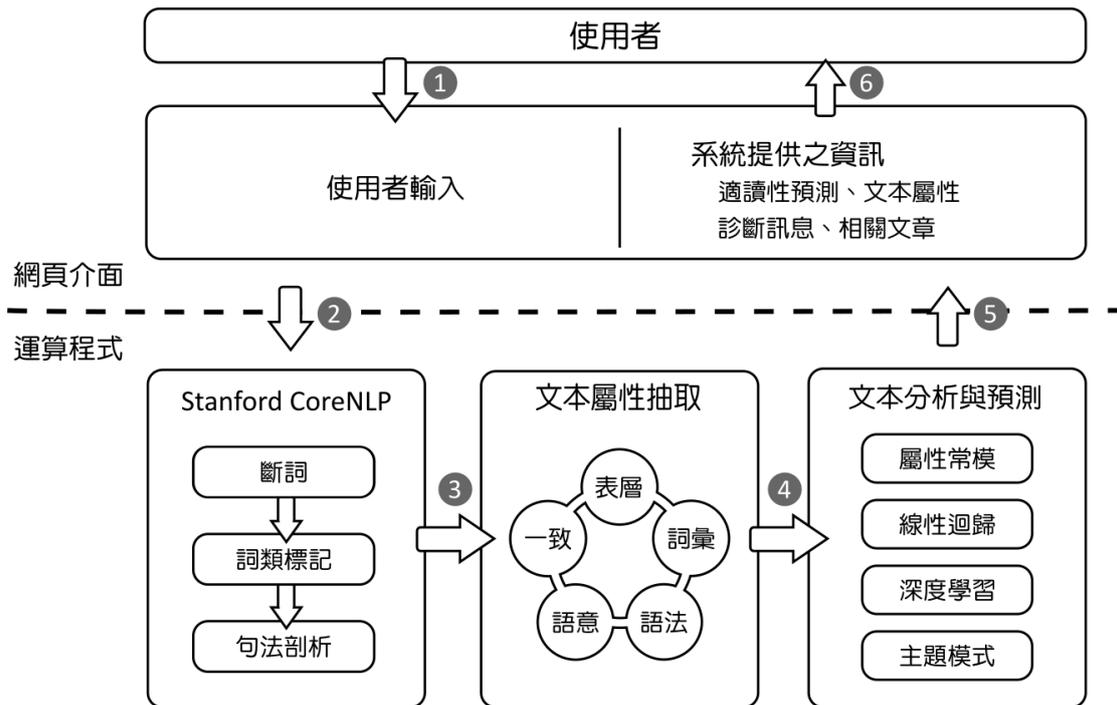


圖 5.1：診斷式系統架構

第二節 系統功能

本系統主要提供四項功能：預測文本適讀年級、分析文本屬性、提供屬性與文本診斷，以及提供相似的文章。此四項功能皆可從系統的主要介面（圖 5.2）點選使用。本系統介面分為三大部份：左側和右側為訊息呈現視窗，分別顯示文本適讀年級、和簡略的文本屬性診斷性提示。主要訊息則呈現在中央主視窗。該視窗有 4 項用途：（1）使用者輸入待分析的文本、（2）呈現文本的屬

性與相對應的常模、(3) 呈現屬性與文本的診斷訊息，與 (4) 提供相似文章。

當使用者輸入文本並按下分析之後，文本的適讀年級預測會自動出現在左側視窗。所有與文本的相關訊息（屬性與診斷訊息等）亦會自動更新，文本屬性及其對應的常模也會出現在中央主視窗的「文本屬性」視窗（圖 5.3）。文本屬性視窗共列出 17 項文本屬性，分別是字數、詞數、字筆畫數、字頻的第 25 百分位數、詞頻的第 25 百分位數、詞彙長度中數、子句長、句長、內容詞功能詞比、詞義數的第三四分位數、內容詞的脈絡重複指標、語義脈絡重複指標、句法數最大深度、罕見語式頻率、名詞修飾詞數、主要動詞前詞數、詞彙豐富度。其中，字數、詞數兩個屬性僅是在系統中呈現予使用者參考的文本屬性，該屬性未用在預測適讀年級的模式中。

為方便使用者判斷是哪些文本屬性影響了文本的適讀性，本系統在中央主視窗中設計了「診斷」功能。點選診斷功能後，系統會首先呈現「難字診斷」，亦即在「文本屬性模式」所預測的適讀年級下，有哪些字屬於較難的字。此系統判斷難字的標準，是該字的字頻低於該年級字頻常模的第 10 百分位數。這些「難字」在診斷系統中將以紅點標示於文本上（圖 5.4）。類似地，使用者可點選右側視窗的「詞彙常用度」進入「難詞診斷」，難詞診斷會以橘色字體標示文本中的「難詞」。難詞的判斷標準為一個詞的詞頻低於該年級詞頻常模的第 10 百分位數（圖 5.5）。在句子層次亦有同樣的診斷訊息，例如點選「名詞修飾詞數」，則進入「難句診斷」，畫面會以橘色底線標示哪些句子含有有較多名詞修飾詞。其判斷方式是以該句子的長度大於該年級常模的第 90 百分位數（圖 5.6）。此診斷系統亦可幫助使用者編輯文本，以改變文本的難度。詳細例子請見本章第三節。

針對輸入的文本，中央主視窗的「相似文章」功能會在社會科的課文中，篩選和輸入文本主題相似的課文。此相似文章的建議來源，是用主題模式估計一篇文本的主題分配，並在社會課文中尋找與之最相近的課文當作相似文章。

例如，當使用者輸入一篇有關鄭成功的文本時，「相似文章」功能會找出在社會課文裡，與之相關的文章。這個分析的結果呈現於圖 5.7。



圖 5.2：診斷式適讀性系統主要介面

註：畫面中關於鄭成功的展示用短文擷取自維基百科「鄭成功」條目（維基百科編輯者，2018）。



圖 5.3：文本屬性畫面



圖 5.4：文本難字診斷



圖 5.5：文本難詞診斷



圖 5.6：文本難句診斷



年級	版本	分數	內文
5	Nan1	0.85	十七世紀中期，明朝滅亡。明朝遺臣鄭成功決定繼續對抗清朝，於是轉而來臺驅逐荷蘭人，開發與經營台灣，作為反清復明的根據地。西元1661年，鄭成功的軍隊從今台南上岸，對荷蘭人展開進攻。隔年二月，荷蘭人投降，(…337…)不穩。1683年，清政府派遣施琅攻打台灣，結束明鄭在臺灣的政權。明鄭統治期間，臺灣從荷西的殖民，轉為漢人政權的社會，不但移入了大量漢人，也帶進了漢文化的生活方式和制度，使漢文化逐漸成為台灣社會的主流。
5	Nan1	0.85	為了殖民與統治的需要，日本積極在台灣推動經濟基礎建設，也著手進行農業改革，讓台灣的經濟有了不同的發展。日本統治台灣後，為了促進經濟發展，於是推動土地調查，以促進買賣與開發；創設台灣第一個現代化的金融機(…406…)台灣本地人不可以單獨成立公司，必須依附在日本人之下；強制低價收購蔗田，大部分的農民只能租地耕種；獨佔蔗糖的交易權，農民的權益完全掌握在日本人手中。這造成大部分民眾無法享受經濟繁榮的成果，生活仍很艱困。
5	Nan1	0.85	清朝末年，西方列強與日本的勢力紛紛進入台灣，改變了清政府治理臺灣的態度，讓台灣開始走向近代化之路。台灣的地理位置，位於西方國家通往中國與日本等國必經的航道，加上盛產茶葉與甘蔗等具有經濟價值的農作物，於(…625…)總局、電報局等，改善台灣內外的運輸與通訊。清查賦稅，讓社會更公平；裝設街燈，建設近代化的台北街道。經由劉銘傳積極的推動各項建設與措施，使台灣各方面進步快速，逐漸邁向近代化，並成為清朝時期最進步的一省。

圖 5.7：與輸入文本相似的文章

第三節 文本診斷與編輯

本系統的功能除了提供各種文本屬性，以提升適讀性預測模式的透明度外，使用者也能藉著系統對文本屬性的分析和提供的診斷訊息，有系統地修改文本，以改變文本的難度。例如以下文本（李琦瑋，2018）：

荒野保護協會昨天公布「二〇一八年臺灣 ICC 淨灘行動數據」，購物用塑膠提袋數量為歷年來最少，證明政府限塑政策有成效，但塑膠吸管卻逐年攀升，連續三年穩坐臺灣海洋廢棄物十大排行榜第三名，呼籲民眾應從源頭減量。

此文本在「深度學習模式」的預測為 6 年級。在「相似文章」的建議中，也出現 5-6 年級的社會與生態議題。「文字屬性模式」的預測則是 5 年級，可能與這篇文章的句子較長、用字用詞較難有關。在診斷系統的文本屬性視窗中，

也可發現與上述觀察一致的訊息（如圖 5.8）：此文章的字彙、詞彙常用度均在常模中低於 5、6 年級的數值，句長也高於六年級的常模數值。

輸入 文本屬性 診斷 相似文章

文本診斷（紅點標示難字）

荒野保護協會昨天公布「二〇一八年臺灣ICC淨灘行動數據」，購物用塑膠提袋數量為歷年來最少，證明政府限塑政策有成效，但塑膠吸管卻逐年攀升，連續三年穩坐臺灣海洋廢棄物十大排行榜第三名，呼籲民眾應從源頭減量。

目標年級

1	2	3	4	5	6
---	---	---	---	---	---

文本屬性

- 詞（字）彙層次
 - 字彙常用度 ● 檢視
 - 字彙筆畫數 ● 檢視
 - 詞彙常用度 ● 正常
 - 詞彙長度 ● 正常
- 子句/片語/句子層次
 - 句子長度 ● 檢視
 - 名詞修飾詞數 ● 正常
 - 句法相似性 ● 正常
- 文本層次

圖 5.8：文本診斷畫面

若使用者想要讓這篇文本變的「容易」一些，他可在診斷系統中選擇目標年級，並點選相關的文本屬性，即可瞭解文本中有哪些可能需要修改的字或詞（如圖 5.7）。從字彙常用度的診斷結果中，使用者可發現「荒野」、「淨灘」、「塑膠」、「海「洋」、「廢棄」物、「呼籲」是較難的字彙。有些字彙牽涉到專有名詞（如「荒野保護協會」）或文本的主要概念（如塑膠）可能會較難替換，但其他詞彙可能可以改成較泛指詞彙，如「呼籲」變成「希望」，「廢棄物」改為「垃圾」等。在句子層次上，這篇短文僅包含一句話，故句子長度較長。使用者可在保留整體行文結構的條件下，將一個句子切成多句敘述，例如：

昨天公布今年清潔海灘的報告，塑膠垃圾是近幾年最少的。這代表大家開始少用塑膠袋是有用的。不過，吸管垃圾卻逐年上升，連續三年都是第三名。大家應該減少使用。

做完上述編輯後重新分析，深度學習模式對新文本的適讀年級預測同樣是 6 年級，但文字屬性模式已變為 3 年級。在這個例子中，雖然使用者無從得知這篇文本「真實」的適讀年級。但使用者可藉著此診斷系統提供的訊息，一步步修改文章中的用字用詞，及修改句子，而達成修改文本難度的目的。

上述例子用一篇文本作為示例，僅是在概念上演示診斷訊息如何幫助使用者編輯文章的難度。在真實教學情境下，文本可能複雜很多，文本中的字、詞、句子各層次的訊息交互影響，都會讓編輯難度的目標更複雜。為了瞭解此診斷系統是否在教學上能真正幫助教師們選擇、編寫教學材料，研究五將進一步實際詢問教師對於此診斷系統的意見。

第六章 研究五、評估診斷系統



研究五之目的是評估研究四所建立的診斷式適讀性系統，是否確實有助於教學現場的小學教師瞭解、調整教材的適讀性。

第一節 研究方法

壹、參與者

此研究之參與者與研究三相同。研究者對參與者的訪談分為兩部分。研究者先進行研究三的訪談與問卷調查，完畢後，再進行本研究的資料收集。

貳、問卷與施測程序

接著研究三的訪談與調查，研究者對參與者進行研究五的問卷調查。問卷開始前，研究者將對參與者說明此問卷將請參與者評估一套此研究自行開發的文本適讀性診斷系統。研究者以口頭介紹系統功能，並以筆記型電腦實際展示系統。系統展示包含介紹系統整體介面，包括（一）參與者可以自行輸入文章，並得到各種適讀性模式的預測結果（如研究四的圖 5.2）；（二）系統會按照適讀性預測的結果，呈現各項文本屬性的診斷訊息。研究者會逐一向參與者展示在診斷系統中，字彙層次、詞彙層次和句子層次的診斷性結果（如研究四的圖 5.4、圖 5.5 與圖 5.6）。研究者介紹後，將讓參與者自由使用此系統的各項功能。

在系統介紹完畢後，參與者將以紙筆問卷形式回答 10 題 5 點量表題目，問卷題目如表 6.1。問卷最後有三題開放式問答题，分別詢問參與者認為此系統在

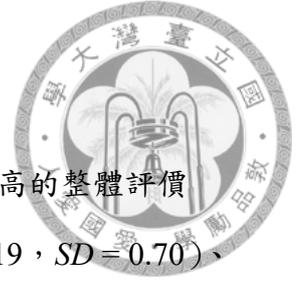
下列部份有哪些可改進之處，包括「整體版面配置」、「呈現文本屬性的方式（例如以紅點標示難字、橘色字標示難詞、底線標示難句等方式）」、「其他可幫助診斷文本適讀性之訊息」。參與者可以紙筆書寫回應，或經參與者同意後，直接以錄音方式記錄參與者的口頭回應。



表 6.1 診斷式系統問卷之題目列表

	很 不 同 意	不 同 意	普 通	同 意	很 同 意
文本診斷					
1. 我認為在字彙層次上，系統標註文本中的難字有助於診斷文本適讀年級。	1	2	3	4	5
2. 我認為在詞彙層次上，系統標註文本中的難詞有助於診斷文本適讀年級。	1	2	3	4	5
3. 我認為在句子層次上，系統標註文本中的困難句子有助於診斷文本適讀年級。	1	2	3	4	5
4. 我認為此系統可按照目標年級顯示需檢視的文本屬性，是有幫助的。	1	2	3	4	5
5. 我認為此系統可幫助我瞭解文章困難的地方	1	2	3	4	5
6. 我認為此系統可幫助我在編寫文章時，調整文章的難度	1	2	3	4	5
整體意見					
1. 我認為此系統整體版面配置清楚	1	2	3	4	5
2. 我認為此系統訊息呈現有條理	1	2	3	4	5
3. 我能在此系統中找到和適讀性相關的文本訊息	1	2	3	4	5
4. 我認為此系統呈現的訊息內容容易理解	1	2	3	4	5

第二節 結果與討論



從問卷結果中發現，參與者對診斷系統的診斷功能有相當高的整體評價 ($M = 4.22$, $SD = 0.50$)。他們認為系統中標註的難字 ($M = 4.19$, $SD = 0.70$)、難詞 ($M = 4.25$, $SD = 0.68$)、難句 ($M = 4.19$, $SD = 0.79$) 有助於診斷文本的適讀年級，同時系統可依照目標年級的不同，呈現不同的診斷訊息，亦有助於瞭解文本的難度 ($M = 4.12$, $SD = 0.81$)。此系統亦可幫助教師瞭解文章困難的地方 ($M = 4.19$, $SD = 0.65$)，以及讓他們在編寫文章時，調整文章難度 ($M = 4.32$, $SD = 0.59$)。

整體系統設計上，參與者亦對此診斷系統的設計有很高的正面評價 ($M = 4.36$, $SD = 0.52$)。他們認為整體版面配置清楚 ($M = 4.39$, $SD = 0.66$)、內容呈現有條理 ($M = 4.52$, $SD = 0.51$)。他們也能在系統中找到和適讀性相關的文本訊息 ($M = 4.16$, $SD = 0.69$)，並認為此系統呈現的訊息容易理解 ($M = 4.39$, $SD = 0.66$)。

在最後開放式問題中，參與者提出的整體期待與建議包括：(一) 此系統應提供更多與詞彙相關的訊息，例如詞彙與年級的對應、或名詞的補充資料與解釋，或甚至提供難詞的替代詞彙選擇等。若有這些訊息更能貼近教師的備課需求。(二) 系統應提供文章的主要宗旨、類型或主題句等較語意的診斷訊息。有些概念對低年級學生無論如何都太難 (例如革命)，故系統在診斷上應能提示文章的語意層面訊息。(三) 此系統應注意多義詞在教學中的重要性，例如「一點」可以指書法中的「一點一捺」，或時間上的「一點」，或指程度的「差一點」，詞彙的不同詞義在教學上，以及學生對於該詞彙概念的掌握程度有很大的差異。(四) 其他對於系統介面設計的整體建議，例如字體大小、字形、顏色顯示等。

整體而言，研究五的結果顯示教師對於診斷系統抱持相當正面的態度，認

為該系統的確可幫助他們在教學時編輯文章教材、及瞭解難度。同時，參與者亦建議諸多未來此系統可發展之方向，例如分析文本內容及主題句、提供困難字詞的代換建議、整合各項詞彙概念的解釋，以及能區辨文本中中文多義詞的不同面向等。這些未來發展皆可讓此系統在教學現場提供更具體的協助。



第七章 綜合討論



此論文的目的為建立一套適讀性的診斷式系統。在五個研究中，研究者首先（一）建立自動抽取文本屬性的自動程序，並以更具體的罕見語式頻率和最大句法樹深度屬性操作過去文獻重視的句法複雜度屬性。（二）開發線性迴歸、支持向量機、主題模式和深度學習四種預測模式。其中，主題模式是以文本主題預測文本適讀性，該預測方式是過去適讀性文獻中較少見的預測取向。（三）比較四種模式的預測表現，並發現整體而言深度學習模式的表現最佳，但四種預測模式彼此間的相關和與文本所在年級的相關皆達顯著水準，顯示4種模式在實用上已具備應用價值。（四）詢問教學現場教師對於四種適讀性預測模式的意見，結果顯示教師較願意採用與文本主題、或與文本屬性（如句法複雜、字詞難度）較相關的主題模式或線性迴歸模式。（五）參考研究三的結果與教師意見設計文本適讀性診斷式系統。該系統除了提供深度學習模式的預測結果，並提供各類文本屬性及相關主題文章，協助教師理解文章難度來源，並可幫助編輯文章難度。（六）再次請教師評估診斷系統。評估結果發現該系統確有助於理解文章難度和編輯文章難度。

研究一與研究二的結果顯示，可發現文本屬性在適讀性的預測上仍然扮演非常重要的角色。首先，在控制文章長度後，迴歸模式可僅用簡單的表層類別屬性（字詞頻、句子長度等）即解釋65%的文本年級變異量。若在迴歸模式加上其他與閱讀心理學相關的文本一致性屬性，及本研究所抽取的句法複雜度屬性後，模型解釋力已接近80%。在正確率上，迴歸模式在國語課文中的表現是四個模式中最佳。這些結果代表國文課本的編撰者在編選文本時，非常重視文本的屬性。當然，因為這樣，採用迴歸分析建構的適讀性模式對於國文文本的適讀性預測也是最有效的。



文本屬性的定義與操作是來自於過去近百年來的適讀性研究所累積下來的知識基礎 (Dale & Chall, 1949; Flesch, 1948; Graesser et al., 2004; Kintsch & Dijk, 1978; Kintsch & Mangalath, 2011)。甚至, 如何描述文本中句子的句法特性, 更需援引過去語言學的研究成果 (Chomsky, 1956; Covington, 2001)。尤其隨著各項自然語言處理技術的進步, 以往無法自動化抽取的文本屬性 (如句法結構等) 已逐漸容易取得。本論文即藉助近年來發展的自然語言處理工具抽取語句中的句法樹結構, 並計算出語式頻率和句法樹深度等屬性。這些屬性更具體的操作過去認知心理學或語言學的理论構念。Domingos (2012) 即指出, 雖然各種機器學習模式能用較多的參數達成預測作業, 但若研究者能仔細的透過領域知識 (domain knowledge) 設計出有用的特徵 (features), 各種分類作業都可大幅簡化。從此研究的結果中, 顯現過去數十年的適讀性研究結果所提供的各類文本屬性能讓研究者預測文本適讀性。

然而, 此研究結果卻也暗示以文本屬性預測適讀性的不足之處。例如, 本研究發現針對「所有文本」或社會文本, 以文本屬性預測社會課文的模式 (線性迴歸、支持向量機) 在預測文本的適讀性上, 皆不如直接以詞彙作為輸入的模式 (主題模式、深度學習)。此現象反應文本類型的不同, 預測模式需著重不同的面向: 國語課文是語言教學導向的文本, 故其文本難度會較反應在文本屬性上; 社會課文則是傳遞知識的文本, 其知識結構較為明確, 故其文本適讀年級可能與其內容或主題較相關。本研究採用主題模式分析社會文本所包含的主題, 並以這些主題分類文本的所在年級。研究結果顯示, 此方式對社會課文的分類可達到 67% 的正確率。然而, 對國文文本的分類, 主題模式的表現不佳。

與其他模式相比, 深度學習雖是教師較不願意採用其結果的預測模式, 但它對社會課文和全部課文的適讀性預測卻是表現是最佳的。而且, 深度學習的神經網絡模式具有多樣的架構選擇, 讓此類模式有很大的改善彈性。本研究的深度學習是一個相對簡單的模式架構, 僅有嵌入層、平均化層、之後即產生

適讀年級預測。本研究模式選擇或許受限於資料量，故未能開發出更完整的模式架構。但近年來的深度學習研究，卻揭示各種深度學習模式的可能。而有些模式架構或許能改善深度學習的透明度，以提升使用者實際採用其預測結果的意願。



有些深度學習模式研究試圖讓模式自動學習到文本語言中的結構屬性。亦即，雖然深度學習模式的輸入資料是文本詞彙，並不包含文句的結構訊息，但只要給予模式適當的架構彈性，模式可自動從詞彙中學習到文句的結構。這些研究使用到循環網絡架構（recurrent neural network），該架構使得模式有足夠的彈性捕捉人類語言的句法特性（Elman, 1990；Hochreiter & Schmidhuber, 1997）。近年來，有些研究更直接分析循環網絡中的表徵，試圖尋找模式學習到哪些和語句結構有關的訊息。例如，Karpathy、Johnson 及 Li（2016）利用長—短期記憶模組（Long-short term memory, LSTM）建立循環神經網絡模式，並讓該模式學習如何預測小說或程式碼中的字彙（character）。當模式訓練完成後，它們分析模組中的每一個單位（cell）的表徵內容，並發現某些 LSTM 中的單位負責表徵語句中的句長、引號、句子位置等，與語句結構有關的訊息。這些研究不僅顯現深度學習模式從詞彙輸入中掌握句法結構的能力，亦顯示至少在特定的作業和架構下，深度學習的表徵並非完全不可穿透的。

另外一些深度學習的研究則試圖描述模式的預測與哪些輸入資料有關。這些研究可讓深度學習模式提供文本的診斷訊息，亦即文本中的哪個字、詞、句子可能是會影響到文本適讀性預測的。此類型的研究來自於神經機器翻譯

（neural machine translation）（Bahdanau, Cho & Bengio, 2015；Luong, Pham, & Manning, 2015; Vasawani et al., 2017）。在機器翻譯模式中，模式需學習原始句和翻譯句之間的關係，但兩句間不必然有相同的詞彙個數，甚至詞彙間也不必然是一對一的對應關係。為了解決這個問題，研究者試圖賦予模式「注意力機制」（attention mechanism），讓模式可學習在產生某個翻譯句中的詞彙時，該

「注意」哪個來源句中的詞彙。在模式中，此注意力是一個可被量化的向量，該向量不僅讓模式得以選擇輸入資料與預測判斷間的權重，研究者亦能透過向量中的權重瞭解每個模式的預測輸出是來自於哪些來源詞彙（Ghader & Monz, 2017）。深度學習的注意力機制或許是能幫助研究者連結輸入資料與適讀性預測之間的未來可能。

本研究建立的診斷系統整合文本屬性和模式預測訊息，並提供文本的診斷訊息，讓使用者能瞭解，可能是哪些文本屬性造成文本的難度。這個系統也可以協助使用者編輯文本，改變文本的難度。從研究三的訪談研究中可知，教師期待適讀性系統能考慮到文本所用字詞的難度、文句的複雜度以及文本所用的主題。這些訊息並非任和一種適讀性預測模式所能提供的。根據研究二的分析，四種適讀性模式雖然預測文本適讀性的正確率不一，但在對各種類型的文本適讀性的預測皆達到顯著相關，所以不同模式對文本分析後，得到的訊息應可以互相整合的。故本研究設計的適讀性診斷系統除了能提供深度學習模式對文本適讀性的預測結果外，還能未受試者提供對瞭解文章難度最有幫助的線性迴歸模式與主題模式對文本分析的結果。同時，系統中也針對文本的字、詞、句子層次，提供各項文本屬性的診斷訊息。這可以讓使用者可具體瞭解到一篇文本的文本屬性影響了文本的難度。最後，為讓使用者瞭解目標文文的主題內容，診斷系統也以主題模式的訊息找出與之相似主題的文章。從研究五的結果評估中，教師們認為此系統確可幫助他們瞭解以及編輯文章難度。

本論文以輸入透明度和參數透明度評估四種不同適讀性模式：線性迴歸、支持向量機、主題模式和深度學習。其中高輸入透明度模式需要使用文本屬性建構適讀性模型，所以在研究一中，研究者建立能夠自動在文本中抽取文本屬性的機制；無論這個文本屬性是傳統適讀性研究使用的表層屬性，或是認知取向模式使用的文章一致性屬性，還是本研究計算之句法複雜度屬性，均能從文本中計算取得。

本論文建立、分析、比較四種適讀性模式的表現，同時以訪談研究詢問小學現場教師最願意採用的適讀性模式。最後，本研究根據研究一至研究三的結果，建立一套診斷式的適讀性系統，該系統可幫助使用者瞭解文章難度和內容主題、以及協助使用者編輯文章難度，而針對小學教師的問卷調查，他們接受這套系統的意願頗高。





第八章 參考資料



- 中文詞知識庫小組（2004）：《中文詞類分析》。台北市：中央研究院。
- 李琦瑋（2018年12月8日）。限塑收效：淨灘塑膠提袋量歷年最少：《國語日報》，第一版。
- 宋曜廷、陳茹玲、李宜憲、查日蘇、曾厚強、林維駿…張國恩（2013）：〈中文文本可讀性探討：指標選取、模式建立與效度驗證〉。《中華心理學刊》，55，75-106。
- 胡志偉、方文熙、李美綾（1994）：〈文章的適讀性指標〉。「第四屆世界華語文教學研討會：語文分析組」，臺北，臺灣。
- 胡志偉、方文熙（1995）：〈中文讀者斷詞的一致性〉。「華心理學家學術研討會暨第三屆中國人的心理與行為科際研討會」，臺北，臺灣。
- 柯華葳、陳明蕾、廖家寧（2005）：〈詞頻、詞彙類型與眼球運動型態：來自篇章閱讀的證據〉。《中華心理學刊》，47，381-398。
- 荊溪昱（1995）：〈中文國文教材的適讀性研究：適讀年級值的推估〉。《教育研究資訊》，3，113-127。
- 馬建忠（1989）：《馬氏文通》。上海：商務印書館。
- 陳世敏（1972）：《中文可讀性公式試擬》（未出版之碩士論文）。政治大學，臺北。
- 程祥徽、田小琳（2015）：《現代漢語》。台北市：書林出版有限公司。
- 黃居仁、謝舒凱、洪嘉馥、陳韻竹、蘇依莉、陳永祥、黃勝偉（2008）：〈中文詞彙網路：跨語言知識處理基礎架構的設計理念與實踐〉。「第九屆漢與詞彙語意學研討會（CLSW 2008）」，新加坡國立大學，新加坡。
- 黃昌寧、趙海（2007）：〈中文分詞十年回顧〉。《中文信息學報》，21，8-



曾昱翔、胡志偉、羅明、呂明蓁、呂菁菁 (2014)：〈從文字屬性檢驗小學國語課本生字之學習順序的恰當性〉。《教育心理學報》，46，251-270。

維基百科編輯者 (2018年5月10日)。鄭成功。擷取自

<https://zh.wikipedia.org/wiki/鄭成功>。

Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67-82.

Baroni, M., Dinu, G., & Kruszewski, G. (2014, June). *Dont count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors*. Paper presented at the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD.

Bahdanau, D., Cho, K., & Bengio, Y. (2015, May). *Neural machine translation by jointly learning to align and translate*. Paper presented at International Conference on Learning Representations 2015, San Deigo, CA.

Benjamin, R. G. (2012). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 24, 63-88.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1798-1828.

Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55, 77-84.

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, 1,

79-132.

Bormuth, J. R. (1969). *Development of readability analyses* (Report No. 7-0052).

Chicago, IL: U.S. Department of Health, Education, and Welfare.

Bransford, J., & Franks, J. (1971). The abstraction of linguistic ideas. *Cognitive*

Psychology, 2, 331–350.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16,

199-231.

Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. (1973). *Methodologies for*

Determining Reading Requirements Military Occupational Specialties (Report

No. HumRRO-TR-73-5). Alexandria, VA: Human Resources Research

Organization.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and

subsequent selection bias in performance evaluation. *Journal of Machine*

Learning Research, 11, 2079–2107.

Chall, J. S., & Dale, E. (1995). *Readability revisited: the new Dale-Chall readability*

formula. Cambridge, MA: Brookline Books.

Chang, P., Galley, M., & Manning, C. (2008, June). *Optimizing Chinese word*

segmentation for machine translation performance. Paper presented at the Third

Workshop on Statistical Machine Translation, Columbus, OH.

Chen, D., & Manning, C. (2014, October). *A fast and accurate dependency parser*

using neural networks. Paper presented at 2014 Conference on Empirical

Methods in Natural Language Processing, Doha, Qatar.

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H.

& Bengio, Y. (2014, October). *Learning phrase representations using RNN*

Encoder-Decoder for statistical machine translation. Paper presented at the 2014



- conference on Empirical Methods in Natural Language Processing, Doha, Qatar.
- Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory*, 2, 113-124.
- Coleman, E. B. (1965). *On understanding prose: Some determiners of its complexity* (Report No. GB-2604). Washington, DC: National Science Foundation.
- Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science*, 315, 464-465.
- Covington, M. A. (2001, March). *A fundamental algorithm for dependency parsing*. Paper presented at 39th Annual ACM Southeast Conference, Athens, GA.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42, 475-493.
- Dale, E. (1931). Evaluating Thorndike's word list. *Educational Research Bulletin*, 10, 451-457.
- Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 27, 37-54.
- Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26, 19-26.
- Danielson, W. A., & Bryan, S. D. (1963). Computer automation of two readability formulas. *Journalism Quarterly*, 39, 201-206.
- Davison, A., & Kantor, R. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly*, 17, 187-209.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Retrieved from

<https://arxiv.org/abs/1810.04805>.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, 25, 104–123

Domingos, P. (2012). A few useful things to know about machine learning.

Communications of the ACM, 55, 78-87.

Dormann, C., Elith, J., Bacher, S., Bachmann, C., Carl, G., Carre, G., ...Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27-46.

Dubay, W. H. (2004). *The principles of readability*. Costa Mesa, CA: Impact Information.

Dufty, D. F., Graesser, A. C., Louwerse, M. M., & McNamara, D. S. (2006). Assigning grade levels to textbooks: Is it just readability? In R. Son (Ed.), *Proceedings of the 28th annual meetings of the cognitive science society* (pp. 1251-1256). Mahwah, NJ: Erlbaum.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.

Farr, J. N., Jenkins, J. J., & Paterson., D. G. (1951). Simplification of the Flesch reading ease formula. *Journal of Applied Psychology*, 35, 333-357.

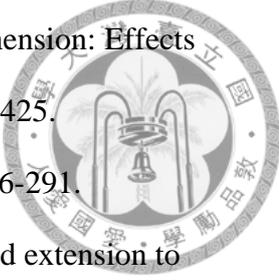
Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In Philological Society (Ed.), *Studies in linguistic analysis* (pp. 1-32). Oxford, England: Blackwell.

Flesch, R. (1943). *Marks of a readable style: A study in adult education*. New York: NY: Columbia University.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.



- 
- Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory Language*, 67, 407-425.
- Fry, E. (2002). Readability versus leveling. *Reading Teacher*, 56, 286-291.
- Fry, E. B. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 21, 242-252.
- Garnham, A. (1987). *Mental models as representations of discourse and text*. Upper Saddle River, NJ: Prentice Hall.
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 245-262.
- Gahl, S., & Yu, A. C. L. (Eds.). (2006). Special theme issue: Exemplar-based models in linguistics. *The Linguistic Review*, 23, 213-379.
- Garnsey, S., Pearlmutter, N., Myers, E., & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58-93
- Ghader, H., & Monz, C. (2017, November). What does attention in neural machine translation pay attention to? Paper presented at The 8th International Joint Conference on Natural Language Processing, Taipei, Taiwan.
- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15, 84-93.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3, 371-398.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223-234.

- 
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36, 193-202.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.
- Grice, P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech Acts* (pp. 41-58). New York, NY: Academic Press.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211-244.
- Gunning, R. (1952). *The technique of clear writing*. New York: McGraw-Hill.
- Habel, C. U. (1982, July). *Referential nets with attributes*. Paper presented at the 9th Conference on Computational Linguistics, Prague, Czechoslovakia.
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349, 261-266.
- Hoosain, R. (1992). Psychological reality of the word in Chinese. *Language Processing in Chinese. Advances in Psychology*, 90, 111-130.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735-1780.
- Hsieh, S. K., Chiang, C. Y., Tseng, Y. H., Wang, B. Y., Chou T. L., & Lee, C. L. (2017). Entrenchment and creativity in Chinese quadrasyllabic idiomatic expressions. In Y. F. Wu, J. F. Hong, & Q. Su (Eds.), *Chinese lexical semantics. CLSW 2017*. Cham, Switzerland: Springer.

Hsieh, S. K., Tseng, Y. H., Lee, C. Y., & Chiang, C. Y. (2018). *Fluid Annotation: A granularity-aware annotation tool for Chinese word fluidity*. Poster session presented at 11th Language Resources and Evaluation Conference, Miyazaki, Japan.



Huang, C. R., Hsieh, S. K., & Chen, K. J. (2017). *Mandarin Chinese words and parts of speech*. Oxford, England: Routledge.

Janssen, N., & Barber, H. (2012). Phrase frequency effects in language production. *PLoS ONE*, 7. doi:10.1371/journal.pone.0033202

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 71-115.

Jurafsky, D., & Martin, H. M. (2008). *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.

Karpathy, A., Johnson, J., & Li, F. F. (2016, May). *Visualizing and understanding recurrent networks*. Paper presented at International conference on Learning Representations 2016, San Juan, Puerto Rico.

Kearns, K. (2011). *Semantics*. London, England: Palgrave.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.

Kintsch, W., & Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, 3, 346-370.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives on memory research*. Hillsdale, NJ: Erlbaum.

- 
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, 133-159.
- Kizilcec, R. (2016, May). *How much information? Effects of transparency on trust in an algorithmic interface*. Paper presented at 2016 Conference on Human Factors in Computing System, San Jose, CA.
- Klare, G. R. (1963). *The measurement of readability*. Ames, IA: Iowa State University Press.
- Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, 10, 62-102.
- Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Literacy Research*, 8, 129-152.
- Klare, G. R. (1988). The formative years. In B. L. Zakaluk & S. J. Samuels (Eds.), *Readability: Its past, present, and future* (pp. 14-35). Newark, DE: International Reading Association.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Redwood City, CA: Stanford University Press.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Liang, S., & Srikant, R. (2017). *Why deep neural networks for function approximation?* Paper presented at 5th International Conference on Learning Representations, Toulon, France.
- Lively, B. A., & Pressey, S. L. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, 9, 389-398.
- Long, D. L., Golding, J. M., & Graesser, A. C. (1992). A test of the on-line status of goal-related inferences. *Journal of Memory and Language*, 31, 634-647.

Luong, M. T., Pham, H., Manning, C. D. (2015, September). *Effective approaches to attention-based neural machine translation*. Paper presented at EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal.



Lorge, I. (1939). Predicting reading difficulty of selections for children. *Elementary English Review*, 16, 229-233.

Ma, W. Y., & Chen, K. J. (2003, July). *A bottom-up merging algorithm for Chinese unknown word extraction*. Paper presented at the Second SIGHAN Workshop on Chinese Language Processing, Sapporo, Japan.

Manning, C. D. (2015). Computational linguistics and deep learning. *Computational Linguistics*, 41, 701-707.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014, June). *The Stanford CoreNLP natural language processing toolkit*. Paper presented at 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD.

McCall, W. A., & Crabbs, L. M. (1961). *Standard test lessons in reading*. New York: Bureau of Publications, Teachers College, Columbia University.

McCarthy, P. M., Dufty, D., Hempelmann, C., Cai, Z., Graesser, A. C., & McNamara, D. S. (2012). Newness and givenness of information: Automated identification in written discourse. In P. M. McCarthy & C. Boonthum (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 457-478). Hershey, PA: IGI Global.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99, 440-466.

McLaughlin, G. H. (1969). SMOG grading - a new readability formula. *Journal of*

Reading, 22, 639-646.



McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York, NY: Cambridge University Press.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1-43.

McNamara, D., Ozuru, Y., & Floyd, R. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4, 229-257.

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P.-W., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly*, 45, 62-92.

Mikolov, T., Sutskever, I. Chen, K., Corrado, G., & Dean, J. (2013, December). *Distributed representations of words and phrases and their compositionality*. Paper presented at the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV.

Miller, I. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 335-354.

Miller, L. R. (1974). Predictive powers of the Dale-Chall and Bormuth readability formulas. *The Journal of Business Communication*, 11, 21-30.

Mullis, I. V. S., & Martin, M. O. (2015). *PIRLS 2016 assessment framework*.

- Chestnut Hill, MA: TIMSS & PIRLS International Study Center.
- Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance*. New York, NY: Student Achievement Partners.
- Ney, H. (1995). On the probabilistic interpretation of neural network classifiers and discriminative training criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 107-119.
- Nivre, J., Marneffe, M., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., ... Zeman, D. (2016, May). *Universal dependencies v1: A multilingual treebank collection*. Paper presented at the 10th International Conference on Language Resources and Evaluation, Portoroz, Slovenia.
- Pearson, P. D. (2013). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *The Journal of Education*, 193, 1-16.
- Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, 11, 357-383.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Malden, MA: Blackwell Publication.
- Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456-463.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: The MIT Press.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A



perspective for research and intervention. *Scientific Studies of Reading*, *11*, 289-312.



Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*, 191-201.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading*. London, England: Psychology Press.

Rehurek, R., & Sojka, P. (2010, May). *Software framework for topic modelling with large corpora*. Paper presented at the seventh International Conference on Language Resources and Evaluation, Valletta, Malta.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?": Explaining the predictions of any classifier. Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA.

Rodd, J., Gaskell, G. & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*, 245-266.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 216-271). Cambridge, MA: MIT Press.

Schriver, K. A. (2000). Readability formulas in the new millennium: What's the use? *ACM Journal of Computer Documentation*, *24*, 105-106.

Schwarm, S. E., & Ostendorf, M. (2005, June). *Reading level assessment using support vector machines and statistical language models*. Paper presented at the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, MI.



Sherman, L. A. (1893). *Analytics of literature: A manual for the objective study of English prose and poetry*. Boston, MA: Ginn.

Shmueli, G. (2010). To explain or to predict. *Statistical Science*, 25, 289-310.

Smith, D. R., Stenner, A. J., Horabin, I., & Smith, M. (1989). The Lexile scale in theory and practice (Report No. NIH-HD-19448). Washington, D.C.: MetaMetrics.

Song, H. A., Kim, B. K., Xuan, T. L., & Lee, S. Y. (2015). Hierarchical feature extraction by multi-layer non-negative matrix factorization network for classification task. *Neurocomputing*, 165, 63–74.

Stevens, K. C. (1980). Readability formulae and the McCall-Crabbs standard test lessons in reading. *The Reading Teacher*, 33, 413-415.

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika*, 61, 509-515.

Sung, Y. T., Chang, T. H., Lin, W. C., Hsieh, K. S., & Chang, K. E. (2016). CRIE: An automated analyzer for Chinese texts. *Behavior Research Methods*, 48, 1238-1251.

Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H., & Chang, K. E. (2015). Constructing and validating readability models: The method of integrating multilevel linguistic features with machine learning. *Behavior Research Methods*, 47, 340-354.

Szalay, T. G. (1965). Validation of the Coleman readability formulas. *Psychological*

Reports, 17, 965-966.

Thronthike, E. L. (1921). *The teacher's word book*. New York: Columbia University.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003, May). *Feature-rich part-of-speech tagging with a cyclic dependency network*. Paper presented at 2003 Human Language Technology Conference, Edmonton, Canada.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: The use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33, 285–318.*

Tseng, H. C., Hung, H. T., Sung, Y. T., & Chen, B. L. (2016, October). *Classification of text readability based on deep neural network and representation learning techniques*. Paper presented at the 2016 Conference on Computational Linguistics and Speech Processing, Tainan, Taiwan.

Tseng, H. S., Chang, P., Andrew, G., Jurafsky, D., & Manning, C. (2005, October). *A conditional random field word segmenter*. Paper presented at the Fourth SIGHAN Workshop on Chinese Language Processing, Jeju Island, Korea.

Vasawani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, ...Polosukhin, I. (2017, December). *Attention is all you need*. Paper presented at 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA.

van Dijk, T. A. (1976). Philosophy of action and theory of narrative. *Poetics, 5, 287-338.*

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Wagner, M. V., & Girasek, D. C. (2003). How readable are child safety seat installation instructions? *Pediatrics, 111, 588-591.*

- 
- Wallach, H. M., Mimno, D., & McCallum, A. (2009, December). *Rethinking LDA: Why priors matter*. Paper presented at the 22nd International Conference on Neural Information Processing Systems, Vancouver, Canada.
- Xue, N., & Shen, L. (2003, July). *Chinese word segmentation as LMR tagging*. Paper presented at the Second SIGHAN Workshop on Chinese Language Processing. Sapporo, Japan.
- Yamashita, T., Yamashita, K., & Kamimura, R. (2007). A stepwise AIC method for variable selection in linear regression. *Communications in Statistics - Theory and Methods*, 36, 2395–2403.
- Yang, S. J. (1970). *A readability formula for Chinese language*. Madison, WI: University of Wisconsin-Madison.
- Zhu, M., Zhang, Y., Chen, W., Zhang, M., & Zhu, J. (2013, August). *Fast and accurate shift-reduce constituent parsing*. Paper presented at the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Oxford, England: Addison-Wesley Press.
- Zwaan, R. A., & Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

第九章 附錄



第一節 語式常模

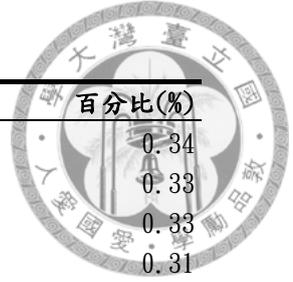
壹、國語課文語式常模

附表 1.1 國語課文語式常模

序號	語式	頻率	百分比(%)
1	一般動詞、名詞	4228	3.63
2	數詞、量詞	2923	2.51
3	名詞、名詞	2701	2.32
4	一般動詞、一般動詞	2662	2.28
5	副詞、一般動詞	1731	1.48
6	副詞、狀態動詞	1223	1.05
7	名詞、的(genitive)	1087	0.93
8	介系詞、名詞	994	0.85
9	狀態動詞、的(comp)	951	0.82
10	名詞、方位詞	924	0.79
11	代名詞、的(genitive)	805	0.69
12	形容詞、的(genitive)	803	0.69
13	名詞、的(genitive)、名詞	797	0.68
14	形容詞、名詞	773	0.66
15	數詞、量詞、名詞	752	0.65
16	代名詞、的(genitive)、名詞	648	0.56
17	一般動詞、的(comp)	635	0.54
18	限定詞、量詞	573	0.49
19	一般動詞、時態標記	557	0.48
20	副詞、一般動詞、名詞	556	0.48
21	專有名詞、名詞	552	0.47
22	名詞、一般動詞	540	0.46
23	一般動詞、一般動詞、名詞	522	0.45
24	限定詞、名詞	478	0.41
25	一般動詞、名詞、名詞	474	0.41
26	狀態動詞、的(comp)、名詞	464	0.40
27	一般動詞、代名詞	451	0.39
28	一般動詞、副詞、一般動詞	425	0.36
29	數詞、名詞	409	0.35
30	形容詞、的(genitive)、名詞	403	0.35

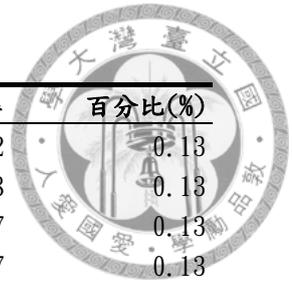
附表 1.1 國語課文語式常模 (續)

序號	語式	頻率	百分比(%)
31	副詞、一般動詞、一般動詞	399	0.34
32	狀態動詞、狀態動詞	388	0.33
33	名詞、連接詞、名詞	381	0.33
34	一般動詞、時態標記、名詞	366	0.31
35	一般動詞、的(comp)、名詞	336	0.29
36	介系詞、名詞、方位詞	331	0.28
37	副詞、副詞、一般動詞	292	0.25
38	有、名詞	285	0.24
39	介系詞、代名詞	279	0.24
40	限定詞、量詞、名詞	278	0.24
41	一般動詞、名詞、一般動詞	266	0.23
42	介系詞、名詞、一般動詞	259	0.22
43	限定詞、數詞、量詞	256	0.22
44	一般動詞、代名詞、的(genitive)、名詞	255	0.22
45	名詞、狀態動詞	245	0.21
46	代名詞、一般動詞	241	0.21
47	一般動詞、名詞、的(comp)	241	0.21
48	專有名詞、的(genitive)	224	0.19
49	代名詞、名詞	220	0.19
50	名詞、名詞、名詞	216	0.19
51	一般動詞、專有名詞	215	0.18
52	名詞、一般動詞、的(comp)	205	0.18
53	狀態動詞、的(manner)	203	0.17
54	一般動詞、名詞、的(genitive)、名詞	201	0.17
55	一般動詞、狀態動詞	193	0.17
56	名詞、名詞、的(genitive)	192	0.16
57	副詞、的(manner)	191	0.16
58	一般動詞、數詞、量詞、名詞	190	0.16
59	一般動詞、名詞、方位詞	186	0.16
60	副詞、狀態動詞、的(comp)	181	0.16
61	介系詞、名詞、名詞	172	0.15
62	一般動詞、一般動詞、一般動詞	169	0.14
63	名詞、一般動詞、名詞	168	0.14
64	一般動詞、一般動詞、的(comp)、名詞	168	0.14
65	名詞、方位詞、的(genitive)	166	0.14
66	副詞、副詞、狀態動詞	165	0.14
67	其他助詞、一般動詞	160	0.14
68	一般動詞、形容詞、名詞	157	0.13
69	一般動詞、名詞、的(comp)、名詞	156	0.13
70	一般動詞、形容詞、的(genitive)、名詞	156	0.13



附表 1.1 國語課文語式常模 (續)

序號	語式	頻率	百分比(%)
71	有、數詞、量詞	152	0.13
72	專有名詞、的(genitive)、名詞	148	0.13
73	繫詞、名詞	147	0.13
74	名詞、名詞、的(genitive)、名詞	147	0.13
75	一般動詞、狀態動詞、的(comp)、名詞	145	0.12
76	副詞、名詞	144	0.12
77	副詞、一般動詞、一般動詞、名詞	143	0.12
78	序數、量詞	143	0.12
79	一般動詞、數詞、量詞	140	0.12
80	一般動詞、副詞	128	0.11
81	介系詞、專有名詞	125	0.11
82	代名詞、一般動詞、的(comp)	122	0.10
83	副詞、一般動詞、時態標記	120	0.10
84	副詞、一般動詞、的(comp)	118	0.10
85	名詞、一般動詞、的(comp)、名詞	117	0.10
86	名詞、副詞、一般動詞	117	0.10
87	名詞、副詞、狀態動詞	115	0.10
88	名詞、狀態動詞、的(comp)	115	0.10
89	介系詞、名詞、的(genitive)、名詞	114	0.10
90	數詞、量詞、形容詞、名詞	114	0.10
91	一般動詞、副詞、狀態動詞	114	0.10
92	時間詞、的(genitive)	109	0.09
93	專有名詞、專有名詞	108	0.09
94	副詞、一般動詞、副詞、一般動詞	107	0.09
95	名詞、方位詞、的(genitive)、名詞	106	0.09
96	一般動詞、代名詞、一般動詞	103	0.09
97	一般動詞、時態標記、數詞、量詞、名詞	103	0.09
98	副詞、有、名詞	102	0.09
99	副詞、狀態動詞、的(comp)、名詞	99	0.08
100	一般動詞、名詞、一般動詞、名詞	98	0.08



貳、社會課文語式常模

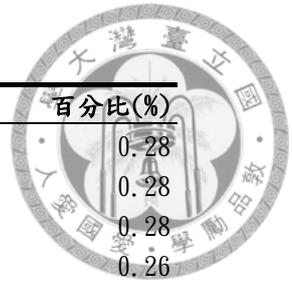


附表 1.2 社會課文語式常模

序號	語式	頻率	百分比(%)
1	名詞、名詞	4484	5.30
2	一般動詞、名詞	2981	3.52
3	一般動詞、一般動詞	1300	1.54
4	一般動詞、名詞、名詞	1167	1.38
5	名詞、的(genitive)	1076	1.27
6	限定詞、名詞	924	1.09
7	副詞、狀態動詞	887	1.05
8	副詞、一般動詞	867	1.02
9	名詞、連接詞、名詞	838	0.99
10	形容詞、名詞	826	0.98
11	介系詞、名詞	670	0.79
12	名詞、的(genitive)、名詞	629	0.74
13	數詞、量詞	577	0.68
14	形容詞、的(genitive)	574	0.68
15	專有名詞、名詞	555	0.66
16	名詞、名詞、名詞	528	0.62
17	名詞、方位詞	471	0.56
18	狀態動詞、的(comp)	461	0.54
19	代名詞、的(genitive)	388	0.46
20	限定詞、量詞	367	0.43
21	副詞、一般動詞、名詞	363	0.43
22	名詞、一般動詞	358	0.42
23	名詞、名詞、的(genitive)	350	0.41
24	一般動詞、一般動詞、名詞	329	0.39
25	數詞、名詞	311	0.37
26	一般動詞、專有名詞	289	0.34
27	一般動詞、的(comp)	283	0.33
28	一般動詞、形容詞、名詞	280	0.33
29	代名詞、的(genitive)、名詞	267	0.32
30	一般動詞、名詞、的(genitive)、名詞	248	0.29

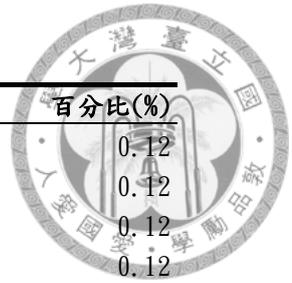
附表 1.2 社會課文語式常模 (續)

序號	語式	頻率	百分比(%)
31	名詞、名詞、的(genitive)、名詞	240	0.28
32	介系詞、名詞、名詞	239	0.28
33	形容詞、的(genitive)、名詞	238	0.28
34	專有名詞、的(genitive)	223	0.26
35	一般動詞、名詞、的(comp)	217	0.26
36	一般動詞、的(comp)、名詞	192	0.23
37	一般動詞、限定詞、名詞	191	0.23
38	一般動詞、副詞、一般動詞	188	0.22
39	名詞、狀態動詞	181	0.21
40	其他助詞、一般動詞	178	0.21
41	副詞、一般動詞、名詞、名詞	177	0.21
42	一般動詞、名詞、連接詞、名詞	175	0.21
43	有、名詞	172	0.20
44	名詞、連接詞、名詞、名詞	171	0.20
45	限定詞、量詞、名詞	170	0.20
46	狀態動詞、的(comp)、名詞	167	0.20
47	名詞、的(genitive)、名詞、名詞	156	0.18
48	介系詞、名詞、一般動詞	156	0.18
49	副詞、狀態動詞、的(comp)	154	0.18
50	一般動詞、名詞、的(comp)、名詞	154	0.18
51	一般動詞、代名詞、的(genitive)、名詞	152	0.18
52	介系詞、名詞、的(genitive)、名詞	151	0.18
53	一般動詞、名詞、一般動詞	151	0.18
54	一般動詞、代名詞	150	0.18
55	一般動詞、名詞、名詞、的(genitive)、名詞	149	0.18
56	副詞、一般動詞、一般動詞	143	0.17
57	一般動詞、名詞、名詞、名詞	140	0.17
58	名詞、一般動詞、的(comp)	140	0.17
59	狀態動詞、狀態動詞	130	0.15
60	一般動詞、形容詞、的(genitive)、名詞	129	0.15
61	介系詞、名詞、方位詞	127	0.15
62	名詞、名詞、連接詞、名詞	120	0.14
63	代名詞、一般動詞	117	0.14
64	形容詞、名詞、名詞	116	0.14
65	一般動詞、一般動詞、的(comp)、名詞	111	0.13
66	一般動詞、一般動詞、名詞、名詞	110	0.13
67	名詞、名詞、連接詞、名詞、名詞	108	0.13
68	形容詞、名詞、的(genitive)	107	0.13
69	一般動詞、狀態動詞	103	0.12
70	名詞、一般動詞、的(comp)、名詞	103	0.12



附表 1.2 社會課文語式常模 (續)

序號	語式	頻率	百分比(%)
71	專有名詞、的(genitive)、名詞	102	0.12
72	一般動詞、專有名詞、名詞	102	0.12
73	狀態動詞、的(comp)、名詞、名詞	101	0.12
74	一般動詞、連接詞、一般動詞	101	0.12
75	專有名詞、名詞、名詞	101	0.12
76	副詞、一般動詞、一般動詞、名詞	100	0.12
77	形容詞、的(genitive)、名詞、名詞	100	0.12
78	介系詞、代名詞	97	0.11
79	名詞、方位詞、的(genitive)	97	0.11
80	介系詞、專有名詞	94	0.11
81	一般動詞、狀態動詞、的(comp)、名詞	93	0.11
82	限定詞、數詞、量詞	92	0.11
83	副詞、狀態動詞、的(comp)、名詞	90	0.11
84	副詞、副詞、狀態動詞	90	0.11
85	限定詞、名詞、的(genitive)	89	0.11
86	副詞、副詞、一般動詞	89	0.11
87	時間詞、的(genitive)	87	0.10
88	數詞、量詞、名詞	87	0.10
89	代名詞、名詞	87	0.10
90	名詞、副詞、狀態動詞	84	0.10
91	專有名詞、名詞、的(genitive)	84	0.10
92	一般動詞、時態標記	83	0.10
93	時間詞、方位詞	83	0.10
94	一般動詞、限定詞、量詞、名詞	80	0.09
95	一般動詞、名詞、的(genitive)、名詞、名詞	79	0.09
96	名詞、名詞、方位詞	79	0.09
97	名詞、連接詞、名詞、的(genitive)	79	0.09
98	有、限定詞、名詞	78	0.09
99	名詞、名詞、一般動詞	78	0.09
100	名詞、一般動詞、名詞	78	0.09



第二節 文本屬性列表



附表 2.1 文本屬性列表

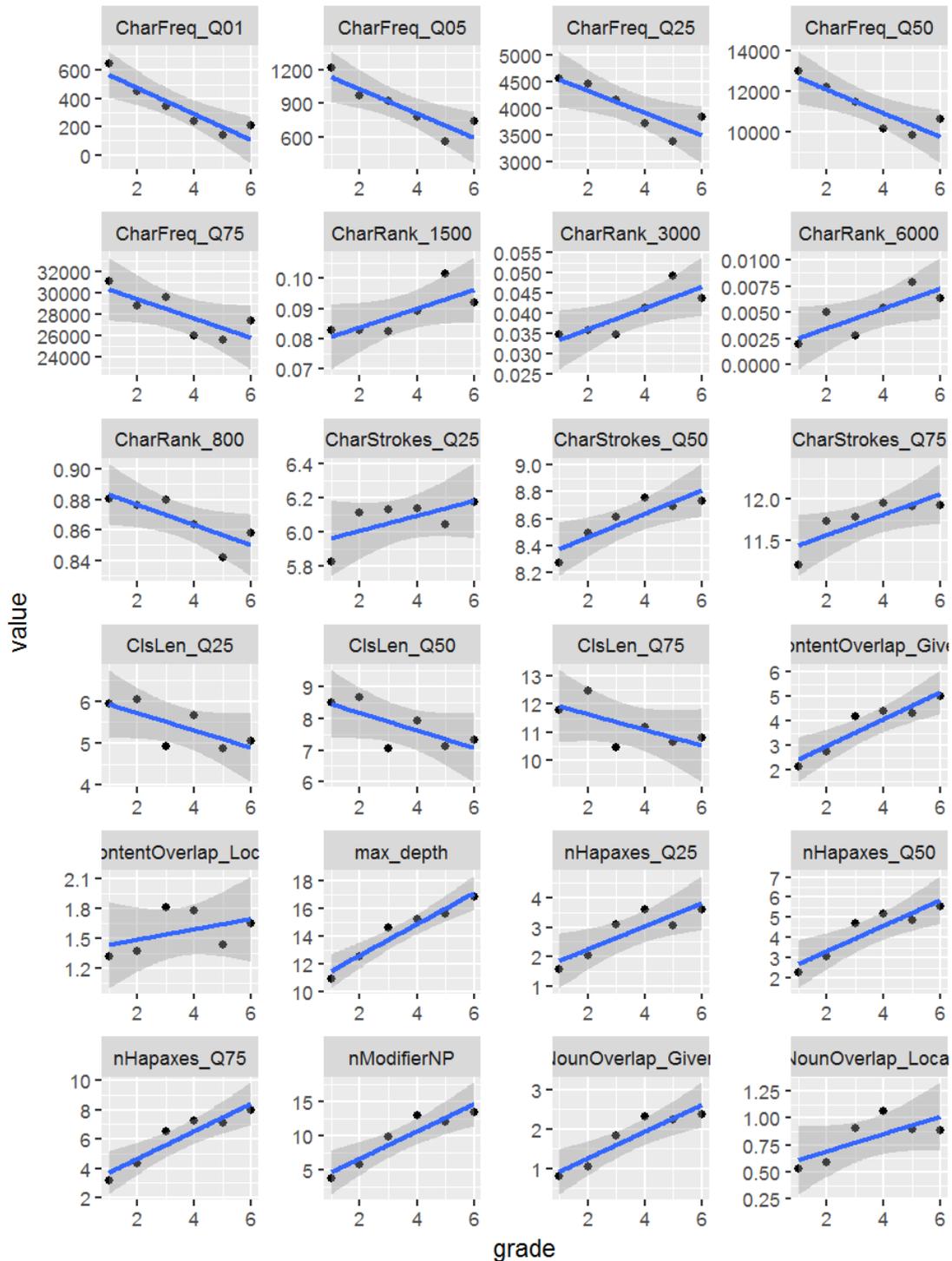
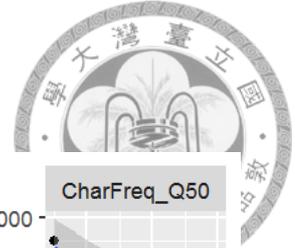
文本屬性	類別	敘述	統計指標
詞 (字) 彙層次			
1 字筆畫數	表層	每個字的筆畫數	Q25, Q50, Q75
2 詞長 (字數)	表層	每個詞所包含的字數	Q25, Q50, Q75 Q01, Q05, Q25, Q50, Q75,
3 字頻	詞彙	字在平衡語料庫中出現的頻率	排序 1-800, 排序 800-1500, 排序 1500-3000, 排序 3000 以上 Q01, Q05, Q25, Q50, Q75,
4 詞頻	詞彙	詞在平衡語料庫中出現的頻率	排序 1-1000, 排序 1000-2000, 排序 5 萬到 10 萬, 排序 10 萬以上
5 名詞比例	詞彙	名詞在文本中出現的比例	比例
6 形容詞比例	詞彙	形容詞在文本中出現的比例	比例
7 代名詞比例	詞彙	代名詞在文本中出現的比例	比例
8 內容詞功能詞比例	詞彙	內容詞相對於功能詞出現比例	比例
9 動詞比例	詞彙	動詞在文本中出現的比例	比例
10 能願動詞比例	詞彙	能、要、會等動詞的出現比例	比例
11 使令動詞比例	詞彙	使、叫、讓等動詞的出現比例	比例
12 連接詞比例	連接詞	文本中出現各類連接詞的總數	比例
13 並列連接詞	連接詞	和、同、與等連接詞出現比例	比例
14 承接連接詞	連接詞	於是、然後等連接詞出現比例	比例
15 遞進連接詞	連接詞	況且、尚且等連接詞出現比例	比例
16 選擇連接詞	連接詞	或者、或是等連接詞出現比例	比例
17 轉折連接詞	連接詞	雖然、雖說等連接詞出現比例	比例
18 因果連接詞	連接詞	因為、由於等連接詞出現比例	比例
19 條件連接詞	連接詞	只有、除非等連接詞出現比例	比例
20 假設連接詞	連接詞	如、若等連接詞出現比例	比例
21 目的連接詞	連接詞	為了、為著等連接詞出現比例	比例
22 解證連接詞	連接詞	例如、譬如等連接詞出現比例	比例
23 詞彙語意複雜度	語意	詞彙在中文詞彙網路的語意數	Q25, Q50, Q75

附表 2.1 文本屬性列表 (續)

文本屬性	類別	敘述	統計指標	
片語層次				
24	子句長 (字數)	表層	每個子句所包含的字數	Q25, Q50, Q75
25	名詞片語比例	句法	在成分語法中名詞片語的比例	比例
26	動詞片語比例	句法	在成分語法中動詞片語的比例	比例
27	介系詞片語比例	句法	在成分語法中介詞片語的比例	比例
句子層次				
28	句長 (字數)	表層	每句所包含的字數	Q25, Q50, Q75
29	把字句	句法	出現「把」字句的比例	比例
30	被字句	句法	出現「被」字句的比例	比例
31	比字句	句法	出現「比」字句的比例	比例
32	名詞修飾詞數	句法	每個名詞前的修飾詞數	平均數
33	主要動詞前詞彙數	句法	每句的主要動詞前的詞彙數	平均數
34	命題深度	句法	以依存語法剖析的句法樹深度	Q25, Q50, Q75
35	句法樹深度	句法	以成分語法剖析的句法樹深度	Q25, Q50, Q75
36	罕見片語語式次數	句法	單一片語語式的出現比例	Q25, Q50, Q75
37	語法相似性	句法	鄰近兩句子的語法相似性	平均數
文本層次				
38	詞類詞例比例	詞彙	文本中詞類數與詞例數的比例	比例
39	句法樹最大深度	句法	句法樹最大深度	最大深度
40	語意局部重複指標	語意	相鄰兩句的詞彙連結度平均	平均數
41	語意脈絡重複指標	語意	每句與前文的詞彙連結度平均	平均數
42	名詞局部重複指標	一致性	相鄰兩句有相同名詞的比例	平均數
43	名詞脈絡重複指標	一致性	每句與前文有相同名詞的比例	平均數
44	內容詞局部重複指標	一致性	相鄰兩句有相同內容詞的比例	平均數
45	內容詞脈絡重複指標	一致性	每句與前文有相同內容詞比例	平均數

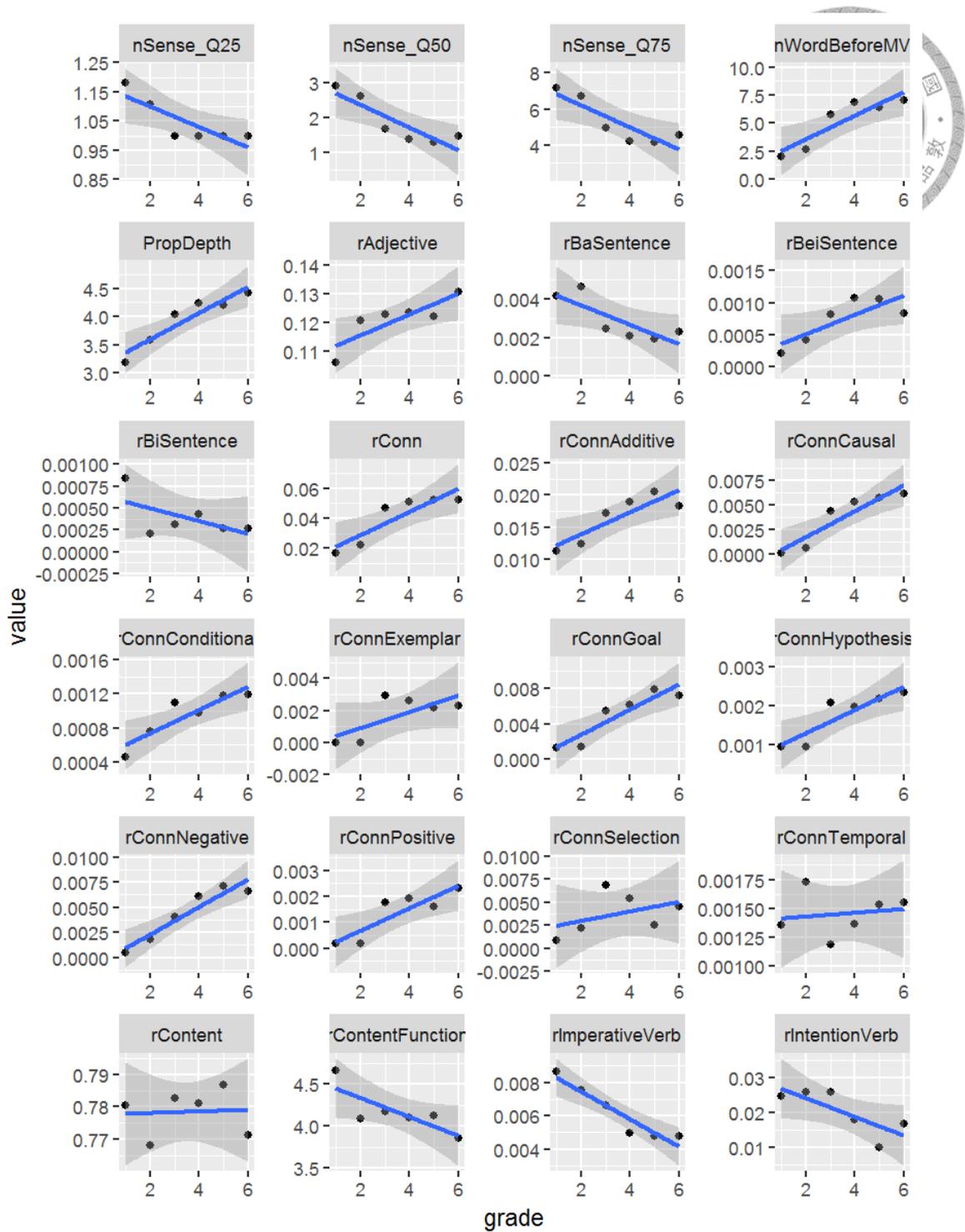


第三節 文本屬性與年級常模

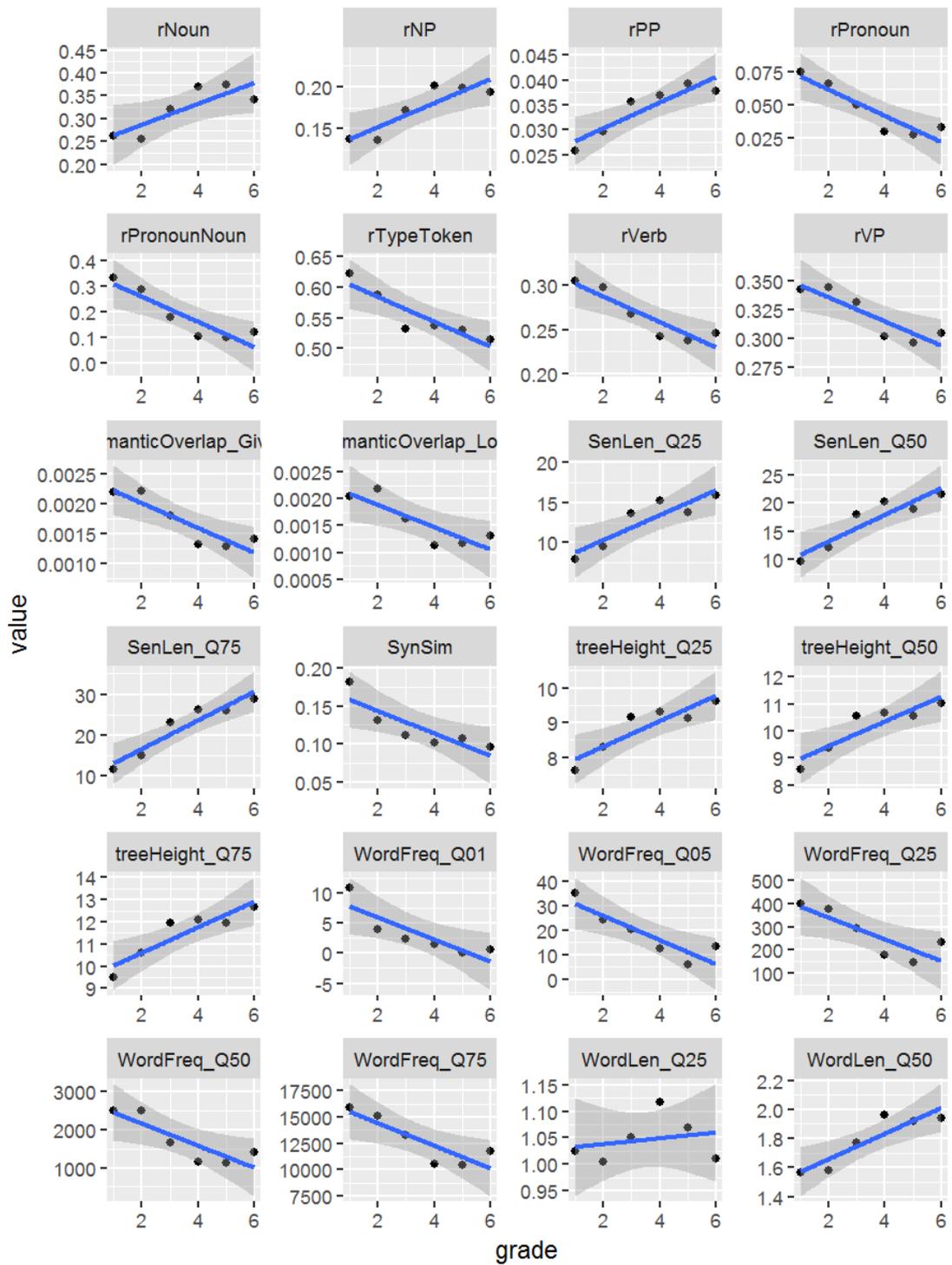
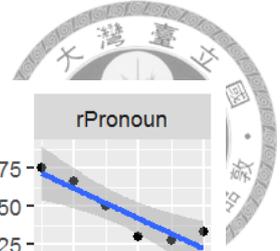


附圖 3.1：文本屬性在國語與社會課本各年級之平均數值（1/4）

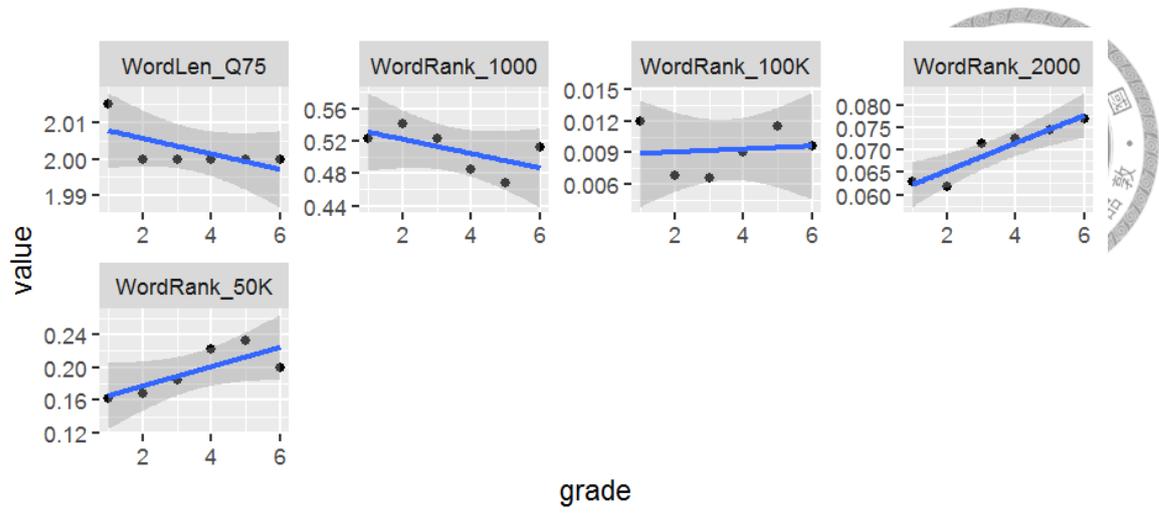
註：圖中趨勢線為各文本屬性在各年級平均值中之線性迴歸線。



附圖 3.2：文本屬性在國語與社會課本各年級之平均數值 (2/4)



附圖 3.3：文本屬性在國語與社會課本各年級之平均數值 (3/4)



附圖 3.4：文本屬性在國語與社會課本各年級之平均數值（4/4）



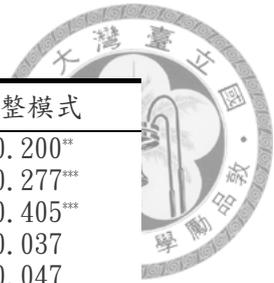
第四節 文本屬性的迴歸分析表

壹、文本屬性與年級的迴歸分析（國語和社會課文）

附表 4.1 文本屬性與年級的迴歸分析（國語和社會課文）

	表層模式	Coh-Metrix 模式	完整模式
詞長 (Q50)	0.130	0.091	0.090
子句長 (Q25)	-0.071***		
詞長 (Q75)			0.034
子句長 (Q50)			-0.041***
句長 (Q25)	-0.245***	-0.132***	-0.093***
句長 (Q75)	0.783***	0.340***	0.257***
字頻 (Q01)	-0.115***		-0.046***
筆畫數 (Q75)		0.033	
字頻 (Q25)	0.110***	0.054***	
字頻 (Q05)			0.072***
字頻 (Q50)	-0.063***	-0.053***	-0.043***
字頻排序 (800)	-0.490	-0.549	-0.531
字頻排序 (1500)	-0.293	-0.326	-0.300
字頻排序 (3000)	-0.194	-0.192	-0.174
詞頻 (Q01)	-0.035***		
詞頻排序 (1000)	0.098	0.128	0.143
詞頻排序 (2000)	0.080	0.060	0.062
詞頻排序 (5 萬)	0.148	0.083	0.086
詞頻排序 (10 萬)	0.036		
名詞比例		0.100	0.134
形容詞比例		0.041	
代名詞比例		-0.177	-0.225
能願動詞比例		-0.040	-0.043
使令動詞比例		-0.032	
詞類詞例比		-0.107	-0.079
內容詞比例		0.047	
內容詞功能詞比		-0.042*	
代名詞名詞比		0.151	0.199
命題深度		0.048	-0.113
動詞片語比例		-0.068	-0.089
連接詞比例		-0.444	-0.447
並列連接詞比例		0.237	0.233
承接連接詞比例		0.056	0.053
遞進連接詞比例		0.065	0.065
選擇連接詞比例		0.069	0.066
轉折連接詞比例		0.149	0.150
因果連接詞比例		0.141	0.132
條件連接詞比例		0.076	0.072
假設連接詞比例		0.092	0.087
目的連接詞比例		0.180	0.169

附表 4.1 文本屬性與年級的迴歸分析 (國語和社會課文) (續)



	表層模式	Coh-Metrix 模式	完整模式
名詞脈絡重複指標		-0.205**	-0.200**
內容詞局部重複指標		-0.321***	-0.277***
內容詞脈絡重複指標		0.461***	0.405***
詞義數 (Q25)		0.036	0.037
詞義數 (Q50)		-0.078	-0.047
詞義數 (Q75)		-0.068**	-0.079***
名詞修飾詞個數		0.103***	0.073***
句法數最大深度			0.214***
罕見語式頻率 (Q75)			0.074**
句法樹高度 (Q25)			0.068
R ²	0.616	0.754	0.771
Adjusted R ²	0.609	0.743	0.760
F Statistic	99.699***	69.521***	74.391***

* $p < .05$. ** $p < .01$. *** $p < .001$.

貳、文本屬性與年級的迴歸分析（國語課文）



附表 4.2 文本屬性與年級的迴歸分析（國語課文）

	表層模式	Coh-Metrix 模式	完整模式
詞長 Q25		-0.083	-0.079
詞長 Q50	0.234	0.137	0.158
詞長 Q75	0.064	0.066	
子句長 Q25	-0.274***	-0.079***	-0.111***
子句長 Q75	0.243***		
句長 Q25	-0.351***	-0.222***	-0.167***
句長 Q50		0.147***	
句長 Q75	0.698***	0.402***	0.374***
筆畫數 Q25	0.088	0.082	0.068
字頻 Q05	-0.125***		
字頻 Q50	0.089***		
字頻排序 800	-0.633	-0.644	-0.617
字頻排序 1500	-0.384	-0.471	-0.416
字頻排序 3000	-0.212	-0.194	-0.192
詞頻 Q01		0.145***	0.066***
詞頻 Q05		-0.086***	
詞頻 Q75	-0.052***	-0.058***	-0.051***
詞頻排序 1000	0.238	0.144	0.134
詞頻排序 2000	0.092	0.061	0.054
詞頻排序 5 萬	0.145	0.121	0.090
詞頻排序 10 萬	0.067	0.086	0.073
使令動詞比例		-0.040	
代名詞比例			-0.081
詞類詞例比		-0.176	-0.161
內容詞比例			0.080
內容詞功能詞比例		-0.052**	-0.058***
代名詞名詞比例			0.098
句法相似度		0.074	0.073
動詞片語比例			-0.101
連接詞比例		-1.133	-1.221
並列連接詞比例		0.600	0.644
承接連接詞比例		0.214	0.237
遞進連接詞比例		0.148	0.150
選擇連接詞比例		0.235	0.243
轉折連接詞比例		0.430	0.457
因果連接詞比例		0.279	0.276
條件連接詞比例		0.217	0.233
假設連接詞比例		0.298	0.306
目的連接詞比例		0.367	0.380



附表 4.2 文本屬性與年級的迴歸分析 (國語課文) (續)

	表層模式	Coh-Metrix 模式	完整模式
名詞脈絡重複指標		-0.217*	-0.203
內容詞局部重複指標		-0.399***	-0.309***
內容詞脈絡重複指標		0.448***	0.332***
詞義數 Q25		0.055	
詞義數 Q50		-0.137*	-0.155*
名詞修飾詞數		0.075***	0.052***
句法數最大深度			0.177***
罕見語式頻率 Q50			0.129**
R ²	0.636	0.773	0.786
Adjusted R ²	0.621	0.751	0.765
F Statistic	43.212***	35.752***	37.625***

* $p < .05$. ** $p < .01$. *** $p < .001$.

參、文本屬性與年級迴歸分析（社會課文）



附表 4.3 文本屬性與年級的迴歸分析（社會課文）

	表層模式	Coh-Metrix 模式	完整模式
詞長 Q25		-0.042	
詞長 Q50	0.105	0.066	0.070
子句長 Q25	0.064***		0.067**
子句長 Q50			-0.067**
句長 Q75	0.632***	0.253***	0.189***
字筆畫數 Q25	-0.071		
字頻 Q01	-0.195***		
字頻 Q25	0.180***	0.095***	0.080***
字頻 Q50	-0.192***	-0.081***	-0.075***
字頻排序 1500			0.056
字頻排序 3000			0.054
詞頻 Q25	0.190***		0.061***
字頻排序 800		-0.071	
詞頻 Q01		-0.039***	-0.043***
詞頻 Q50		0.091***	0.104***
詞頻排序 2000	0.133	0.045	0.044
詞頻排序 5 萬	0.214		
詞頻排序 10 萬	0.044		
動詞比例		-0.183	-0.109
代名詞比例		-0.281	-0.139
詞類詞例比		-0.098	-0.111
功能詞比例		0.076	0.094
代名詞名詞比例		0.143	
主要動詞前詞數		0.083***	0.072***
名詞片語比例			-0.078
動詞片語比例			-0.120
被字句比例		0.034	0.037
連接詞比例		-0.234	-0.219
並列連接詞比例		0.098	0.096
遞進連接詞比例		0.038	0.049
轉折連接詞比例		0.062	0.062
因果連接詞比例		0.084	0.071
條件連接詞比例		0.064	0.065
假設連接詞比例		0.038	0.034
目的連接詞比例		0.106	0.106

附表 4.3 文本屬性與年級的迴歸分析 (社會課文) (續)

	表層模式	Coh-Metrix 模式	完整模式
名詞局部重複指標		-0.102	-0.167
名詞脈絡重複指標			0.216*
內容詞局部重複指標		-0.205*	-0.138
內容詞脈絡重複指標		0.250***	
詞義數 Q25		0.080	0.066
詞義數 Q75		-0.114***	-0.107***
語意局部重複指標		-0.077	-0.084
名詞修飾詞數		0.089***	0.061***
句法數最大深度			0.162***
罕見語式頻率 Q50			-0.092*
罕見語式頻率 Q75			0.120***
句法樹高度 Q25			0.050
R ²	0.678	0.805	0.820
Adjusted R ²	0.671	0.792	0.805
F Statistic	95.713**	63.670**	55.031**

* $p < .05$. ** $p < .01$. *** $p < .001$.

曾昱翔

seantyh@gmail.com



研究興趣

文本分析、自然語言處理、語言與機器學習、文本閱讀

學歷

臺灣大學心理學系學士

臺灣大學心理學系碩士

臺灣大學心理學系博士

經歷

臺大心理學系兼任研究助理、普通心理學教學助理

臺大醫院眼科部兼任研究助理

臺大語言所兼任研究助理

臺大課程「R 語言與資料科學導論」、「人文學群的 Python 入門」教學助理

著作

Kao, C. H., Hue, C. W., **Tseng, Y. H.**, & Lo, M. (2009). Detection errors in proofreading Chinese texts: A study of Chinese reading units and word inferiority effect in detecting character-position transposed words. *Chinese Journal of Psychology*, 51, 21-36.

Hue, C. W., **Tseng, Y. H.**, Lu, C. C., Kao, C. H., Chen, Y. F., & Kou, Y. W. (2009). *Chinese word inferiority effect: A test of Kao's model*. Paper presented at the 19th International Symposium on Theoretical & Applied Linguistics. Thessaloniki, Greece. (oral presented by Tseng, Y. H.)

Tseng, Y. H., Lo, M., & Hue, C. W. (2010). *An economical and practical eye tracking toolbox*. Paper presented at 46th Conference of Taiwan Psychological Association, Tainan, Taiwan.

Hue, C. W., Lo, M., **Tseng, Y. H.** (2011). *Development of phonetic awareness: a test for a statistical approach of language learning*. Paper presented at The Japanese Society for Language Sciences 2011, Osaka, Japan.

Lo, M., Hue, C. W., **Tseng, Y. H.** (2012). C-CAT2: An analysis toolkit of character

- components and neighbors for traditional and simplified Chinese. *Chinese Journal of Psychology*, 54, 243-252.
- Hue, C. W. & **Tseng, Y. H.** (2012). *Nominal and functional semantic functions of Chinese classifiers*. Paper presented at the 14th International Conference on the Processing of East Asian Languages, Nagoya, Japan. (oral presented by Tseng, Y. H.)
- Tseng, Y. H.**, Chu, C. L., Lu, C. C., Lu, M. C., & Hue, C. W. (2012). Words and characters usage frequency in textbooks of elementary school: an analysis of concepts in textbooks. Paper presented at 50th Conference of Taiwan Psychological Association, Taichung, Taiwan.
- Hue, C. W., **Tseng, Y. H.**, Lo, M., Lu, M., Lu, C. C. (2013). *Chinese characters elementary school children learn: which and when*. Paper presented at The 4th International Conference on Han Characters Education and Research, Hong Kong. (oral presented by Tseng, Y. H.)
- Tseng, Y. H.**, Hue, C. W., M. Lo, M. Lu & C. C. Lu. (2014). Examining the appropriateness of teaching sequence of Chinese characters in Chinese elementary school textbooks. *Bulletin of Educational Psychology*, 46, 251-270.
- Lo, M., Hue, C. W., **Tseng, Y. H.**, Chi, T. T., Liaw, M. L., Lu, C. C., & Lu, M. (2014). A survey of the competence that Taiwanese students need to acquire in 12-Year compulsory education. *Educators and Professional Development*, 31, 49-58.
- Tsai, T. H., **Tseng, Y. H.**, Chen, C. C., & Hu, F. R. (2015). *Comparison of interocular suppression of intermittent exotropia and non-strabismic amblyopia with functional magnetic resonance imaging*. Paper presented in Symposium for Integrated Imaging of Body, Mind and Culture & 2015 Forum of MOST Research Projects on Brain Imaging for Mind Science, Taipei, Taiwan.
- Tsai, T. H., **Tseng, Y. H.**, Chen, C. C., & Hu, F. R. (2016). *Interocular suppression in striate and extrastriate cortex in human intermittent exotropia*. Paper presented in the Asia-Pacific Academy of Ophthalmology Congress, Taipei, Taiwan.
- Hsieh, S. K., Chiang, C. Y., **Tseng, Y. H.**, Wang, B. Y., Chou T. L., & Lee, C. L. (2017). Entrenchment and creativity in Chinese quadrasyllabic idiomatic expressions. In Y. F. Wu, J. F. Hong, & Q. Su. (Eds), *Chinese Lexical Semantics. CLSW 2017*. Springer, Cham.
- Hsieh, S. K., **Tseng, Y. H.**, Lee, C. Y., & Chiang, C. Y. (2018). *Fluid Annotation: A granularity-aware annotation tool for Chinese word fluidity*. Paper presented at Language Resources and Evaluation Conference, Miyazaki, Japan.