The 2016 Conference on Computational Linguistics and Speech Processing
ROCLING 2016, pp. 342-351

中文近義詞的偵測與判別*
Detection and Discrimination of Chinese Near-synonyms

李詩敏†Shih-Min Li、白明弘 Ming-Hong Bai、吳鑑城 Jian-Cheng Wu、
黃淑齡 Shu-Ling Huang、林慶隆 Ching-Lung Lin
國家教育研究院編譯發展中心
{smli, mhbai, wujc, slhuang, cllin}@mail.naer.edu.tw

摘要

本文嘗試以自然語言處理及語言學分析之方式，結合語料庫及電子資源，採用中文語料庫、雙語語料庫及 Word2Vec 模型工具及廣義知網電子資訊，以偵測並判別中文近義詞。藉由訓練資料，建立近義詞偵測及判別規則，並利用測試資料評估規則的正確性及可行性；本文建議透過國家教育研究院近義詞系統自動偵測近義詞組，並採用詞性訊息、詞性異同、英文翻譯、詞素數目、雙語對譯等規則，可篩選出豐富的近義詞組。經由本文所建立近義詞系統以及人工判別近義詞規則，可避免過去研究因傳統人工編纂近義詞而費時耗力或是單採中文語料庫而有資料雜訊較多之缺點，同時本文篩選的近義詞具客觀性及正確性，並兼顧詞彙的句法和語用特點，未來可應用至發展中文近義詞自動辨識技術及系統，以及近義詞辨析、辭典編纂、中文教學及文章寫作等方面。

關鍵詞：近義詞，華語文語料庫，Word2Vec，雙語語料庫，詞性

一、前言及目的

探究近義詞，首先必須釐清近義詞的範圍與定義。過去探討中文詞彙語意的研究，部分主張應區分同義詞和近義詞，此派尤以中國大陸學者居多 [1]；對於可替換性是否為同義詞的必要條件，學者們看法不盡相同 [2][3][4]。相較之下，採西方語言學觀點者多認為同義詞少之又少，故以近義詞指稱語意相似的詞彙群組群；例如，Pinker [5] 認為同形異義詞很多，但同義詞很少，事實上所有被認定是同義詞的詞彙語意都仍有差異；Taylor [6][7] 提到，完全同義詞(perfect synonyms)應語意相同且可在任何語境中互相替換，因此完全同義詞的數量極少，大多數同義詞其實是近義詞，這些近義詞共享某些核

* 本文感謝科技部補助延攬科技人才計畫（MOST 105-2811-H-656-001、MOST 105-2811-H-656-002）之 補助與支持。文中若有疏漏及錯誤，概由作者負責。
† 本文通訊作者。

心（語意）相似，而邊緣（語意）相異；Saeed [8] 認為真正的同義詞極少。Chung 與 Ahrens [9] 綜合前人研究指出，同義詞的意涵(connotations)、蘊含(implications)、選擇限制(selectional restrictions)及句法變異(syntactic variations)等各方面均可能有所差異。概括上述研究可知，真正的同義詞數量鮮少，因此本文以近義詞統稱核心語意相同但次要或邊緣語意具細微差異的詞彙群組，並以近義詞作為本文討論範圍。

根據周荐之研究結果顯示 [1]，研究近義詞不可忽視以下現象：不少被認定為近義詞的詞 語是帶有主觀性及隨意性，沒有多少科學依據，甚至不被大家認同，例如以下的近義詞 組：「痛快」和「喜歡」、「忽視」和「蔑視」、「助手」和「幫凶」、「心力交瘁」和「鞠 躬盡瘁」。此外，以人工分析近義詞，雖可藉助語料庫觀察詞彙出現的情形、分佈及頻 率，達到細緻分析，但逐個檢視判斷詞彙關係是否為近義詞的方法極為費工耗時。為避 免上述缺失，本文擬建立綜合自然語言處理技術及語言學知識，先以模型工具自動計算 詞彙的相似度，再透過人工分析找出協助電腦判別近義詞的規則；研究目的主要為利用 自然語言處理之模型工具計算詞彙在真實語境中的相似度，並輔以人工分析建立近義詞 判別規則，客觀且正確提供較遠傳統辭典數量更為豐富的近義詞，此規則未來可應用於 建立中文近義詞自動辨識技術及發展自動辨識系統。

二、文獻探討

鑑於人工選取近義詞過於主觀，也為提升近義詞研究的效能性及科學性，不少研究以計算語言學角度採用電子資源、語料庫等方式，利用程式自動且大量計算詞彙語意的相似度，並將研究成果應用至機器翻譯、文本分類、訊息檢索及解除岐義等方面。採用電子資源計算詞彙相似度者，利用 WordNet（詞彙網路）[10][11]、《同義詞詞林》[12]、知網(HowNet) [13] 等原已建立之電子辭典或語意網路等資源，建構詞彙與詞彙的關係，進而計算出詞彙之間的語意相似度。採用語料庫計算詞彙相似度者，將語意特徵以在語意空間向量的方式表達 [14]，詞彙的相似即是其在向量空間位置的相似，詞彙相似度通常以兩兩詞彙之向量夾角 consine 值來表達 [15][16]，例如應用「背景向量模型」(Context Vector Model) [17]、「潛在語意分析」(Latent Semantic Analysis) [18][19]、Word2Vec [20][21]。

[1] 以下敘述請參照 [1]。

根據劉群、李素建的論述 [2]，採用電子辭典或語意網路等電子資源，雖無需語料庫，仍 可準確反映詞彙語意的相似相異，但容易流於直觀，無法客觀反映事實，且對詞彙句法 和語用特點考量較少；而採用語料庫方法的優點則是客觀，可綜合反映詞彙在句法、語 意及語用的相似及相異，但缺點是計算量大且受資料雜訊(noise)干擾較大。

再比較以上應用語料庫者所採用的技術，「背景向量模型」雖具演算公式，但程式碼不公開，外界無法取得及應用；「潛在語意分析」的程式碼開放，但其處理大量語料時，矩陣會變得龐大，演算時間增加，且「潛在語意分析」對句子線性結構的訊息處理較少；Word2Vec 程式碼開放，處理大量語料的速度快，且它是雙層的類神經網路，將詞彙的上下文語境以矩陣方式表達，可以比較精確找到詞彙間的語意和句法關係，並可以用向量加減法方式調整模型結果以找出最符合需求的詞彙語意關係。

根據蔡美智的研究結果 [22] 顯示，近義詞的語意及語法功能、搭配關係也經常部分相同。鑑於語料庫方法可反映近義詞在語言各層面的相似及相異，其缺點為資料雜訊干擾較大，但此方法遠較採用電子資源更為客觀；因此，為避免單採語料庫之缺點，本文將採用國家教育研究院「華語文語料庫」[3] 大型語料庫之書面語語料 [23]，並利用雙語語 料庫的中英對譯以降低雜訊較大的缺點；再以 Word2Vec 模型為工具，可快速處理大量 語料及自動計算詞彙之間的相似度。

三、研究方法

鑑於「華語文語料庫」資料龐大但斷詞和詞類標記均為電腦自動處理，斷詞及詞類正確性仍有提升空間，相較之下，「中央研究院漢語平衡語料庫 4.0 版」語料雖較少，但斷詞和詞類標記均經過人工檢視，正確性極高，有助於提高近義詞偵測的正確率，因此本文同時採用「華語文語料庫」和「中央研究院漢語平衡語料庫 4.0 版」以達到語料量大及正確性高之雙重優點。[4]

---

[2] 相關論述詳見 [13]。

[3] 「華語文語料庫」（Corpus of Contemporary Taiwanese Mandarin，簡稱 COCT）為國家教育研究院華語 文八年計畫「建置應用語料庫及標準體系」所建置；截至目前為止，共收錄書面語語料 1 億 1,220 萬字，口語語料 960 萬字，華英雙語語料 340 萬字，華語中介語 42 萬字。

[4] 原訂以「華語文語料庫」為本，但在研究過程發現「華語文語料庫」自動斷詞及標記之部分結果降低詞彙比對之正確率。例如，「中心」在「華語文語料庫」全被標記為 Nc（地方詞），無 Na（普通名詞）；然而，坊間辭典將「核心」與「中心」視為近義詞，可知此「中心」指「事物的主體」之意（詞性 Na），而非「正中央位置」或「樞紐地位之地點」（詞性 Nc）。由於「中心」僅被標為 Nc，導致模組工具無法

本文擬結合自然語言處理及語言學之研究方法，先以 Word2Vec 模型為工具，從而計算詞彙在「華語文語料庫」及「中央研究院漢語平衡語料庫 4.0 版」之相似度，以電腦自動偵測出個別詞彙的近義詞群，而後輔以國家教育研究院、臺灣師範大學的雙語索引典等工具，再以人工檢視方式分析訓練資料，從語言學分析角度歸納得到影響詞彙相似度的因素並建立判別近義詞的規則，最後再利用測試資料驗證以上規則，用以評估本文選取近義詞的正確性及完整性。

近義詞資料來自坊間六部辭典（《現代漢語同義詞詞典》（朱景松主編）、《1700 對近義詞語用法對比》、《同義詞詞林》、《現代漢語同義詞典》（賀國偉主編）、《繪圖同義詞辨析造句詞典》、《教育部國語辭典簡編本》）所提供的近義詞資料。例如，在《現代漢語同義詞詞典》，主要詞項「哀求」的同義詞項為「懇求」；在《現代漢語同義詞典》，主要詞項「尊崇」的同義詞項為「尊敬」、「尊重」、「崇敬」、「崇尚」、「敬重」。當某個主要詞項的近義詞項同時被過半數辭典所編納，則此組近義詞即列入本文觀察；換言之，假設「懇求」在三部以上辭典都被認為是「哀求」的近義詞，則「哀求」和「懇求」就列入觀察之列。

從以上列入觀察的近義詞組資料中，從中挑選一部分作為訓練資料，用以分析並建立近義詞偵測及判別規則；再挑選另一部分近義詞組資料作為測試語料，用以確認以上規則的信效度。藉此，預計本文將建立中文近義詞自動辨識技術。

四、結果與討論

以下研究結果分訓練資料及規則建立、測試資料及驗證結果兩大部分來討論。

（一）訓練資料及規則建立

以下以同時出現在六部辭典的「高興」和「愉快」以及同時出現在五部辭典的「根本」和「基本」為例，作為本文判別及建立近義詞規則的來源之一。

利用 Word2Vec 檢索得到的原始語料，「愉快」位於「高興」的近義詞組第 48 名。首先

判讀「核心」（詞性　Na）與「中心」（詞性　Na）的近似值。因此，本文增加已經人工校正之「中央研究 院漢語平衡語料庫4.0版」，作為近義詞比對來源之一。

利用廣義知網(E-HowNet)知識本體架構[5]檢視兩者的語意分類及定義是否相近，但結果不甚理想，兩者的分類層次並不同，「高興」在「MentalState|精神狀態」之下，「愉快」在「AttributeValue|屬性值」之下。接著，因為近義詞的詞性和語法功能接近[6]，本文將詞性特徵[7]加入近義詞判別規則中，其用意有二：一是倘若近義詞本身為多義或多詞性，可利用詞性訊息區別語意和語法行為[8]；二是倘若近義詞本身非多義或多詞類，藉由詞性訊息可增加獲知其鄰近詞詞性的訊息，可協助更進一步細緻分類和區辨，以增強鄰近詞彙鑑別度。加入詞性訊息後，「愉快」位於「高興」的近義詞組排名降為第 77 名，顯示詞性訊息對此組近義詞作用不大。[9]而後，以國家教育研究院、臺灣師範大學的雙語語料庫之中英對譯為輔助工具，利用雙語索引典查詢「高興」的英文翻譯有：*glad*, *happy*, *pleased*, *excited*, *delighted*, *joy*, *enjoyed*，再找出「高興」77 名近義詞候選名單（即加詞性訊息後所偵測出的近義詞結果）的英文翻譯，倘若候選者的英文翻譯為「高興」的英文翻譯其中之一，則候選者被判定為「高興」的近義詞；藉由以上方法，「愉快」排名提升至第 10 名（前 10 名為：「開心」、「興奮」、「興高采烈」、「雀躍」、「歡喜」、「慶幸」、「快樂」、「欣慰」、「快活」、「愉快」）。並非所有學者都認同近義詞的完全可替換性，但無法否定近義詞具部分替換性；本文建議嚴謹的中文詞彙替換及使用應考量構詞的詞素數，亦即，近義詞之間的詞素數目應相同；在本文所參考的六部辭典中，僅《同義詞詞林》未將近義詞組詞素數相同考慮在內，由此可知，詞素數相同仍列為近義詞辭典編纂考量之一；因此，考慮詞素數後，「愉快」的排名變成第 9 名（即刪除「興高采烈」）。

以上本文所偵測及判別的「高興」9 個近義詞，「開心」、「興奮」、「愉快」被五部辭典[10]列為「高興」的近義詞，另一方面，本文模組工具偵測到五部辭典所列的「高興」全部 的近義詞，僅未偵測到《現代漢語同義詞典》所列的「喜歡」。

為了找出更多的近義詞，以「高興」前 9 名近義詞的英文翻譯為本，再利用雙語語料庫找出這些英文翻譯的中文對譯詞彙；為了避免結果過於發散，再取中文對譯詞彙交集數

[5] 2.0 版線上查詢系統網址為 http://ehownet.iis.sinica.edu.tw/index.php。

[6] 相關論述請參考 [1] 及 [26]。

[7] 本文詞性標記參照「中央研究院漢語平衡語料庫」詞性標記 http://asbc.iis.sinica.edu.tw/images/98-04.pdf。

[8] 當詞彙具歧義時，其語意會比較模糊。用意一即在解歧，則訓練模組可降低雜訊問題。

[9] 此時自動偵測的近義詞包括：「難過」、「著急」、「生氣」等，人類知識判別為非近義詞。本文曾考慮增加語意角色訊息，但結果顯示語意角色相同與否對近義詞偵測幫助不大，推測是因為近義詞的語法功能並非完全相同而有此結果。基於同樣原因，語意角色詞組是否同樣為動詞組或名詞組也無助近義詞判別。

[10] 《同義詞詞林》未考慮近義詞組詞素數目，且許多近義詞組語意已偏離原主詞項語意，故本文分析及 驗證時均暫未將此部辭典的近義詞納入評估。類似情形也出現在廣義知網的同義詞/近義詞。

較高者，便找出更多近義詞。結果顯示，「幸福」、「喜悅」、「歡樂」、「滿意」、「樂意」、「美滿」、「欣喜」、「愉悅」、「和樂」、「美好」、「喜歡」、「激動」、「驚喜」、「歡娛」均被判定為「高興」的近義詞。在前一個步驟以近義詞英文彼此對應而找出的近義詞獨漏「喜歡」，利用近義詞英文對應中文的方式擴大尋找「高興」的近義詞，則將「喜歡」納入。

利用以上所歸納的近義詞偵測及判別原則，再檢視「根本」和「基本」的關係，以確定以上所提出的規則具可操作性。由於「根本」和「基本」均有歧義，詞性訊息便發揮效用；首先，在廣義知網，「根本」和「基本」名詞詞性的分類，以及「根本」動詞和「基本」非謂形容詞的分類，兩者均不在同一類別；[11] 原始語料不具詞性訊息，「基本」位於「根本」的近義詞組第 101 名，因為缺乏詞性訊息，故無法得知「基本」身份為何。增加詞性訊息後，名詞「根本」的近義詞排名第 20 名為名詞「基本」，第 28 名為非謂形容詞「基本」；而後剔除和名詞「根本」詞性不同者（包括動詞、非謂形容詞），名詞「基本」排名變成第 17 名。[12] 接著，以雙語語料庫輔助，「根本」的英文翻譯為：*root*, *key*，囿於雙語語料庫字數及資料的侷限，無法以以上英文翻譯找到「基本」；雖然前面提到廣義知網的同義詞/近義詞未列為本文評估工具之一，但其詞彙的英文意涵仍具功效，利用「根本」的英文意涵 *essence*, *foundation* 及「根本」的英文翻譯對應至中文，偵測到更多的「根本」近義詞：「根源」、「根基」、「基礎」、「地基」、「基石」、「基底」、「基本」、「利基」，除了「基本」之外，這些近義詞均未列在五部辭典中，但根據語感判斷，它們也和「根本」享有同樣的核心語意，仍可視為「根本」的近義詞。

經由以上訓練資料的分析，本文建立近義詞偵測及判別的規則如下：

1. 詞性訊息：以達到解歧及增加詞彙鑑別度；

2. 詞性異同：剔除詞性不同（名詞 Na, Nb, Nc, Nd 各為不同分類，動詞 V 和 A 視為一類），以達到近義詞的部分可替換性；

3. 英文（雙語語料庫及廣義知網）翻譯：具部分相同英文翻譯，以達到近義詞的準確率；

4. 詞素數目：近義詞之間的詞素數相同，以達到較為嚴謹的近義詞判定結果；

---

[11] 名詞「根本」分類在「thing|萬物」，「基本」在「Other(object|物體)」；動詞「根本」在「PropertyValue|性質值」，非謂形容詞「基本」在「SituationValue|狀況值」。

[12] 增加詞性訊息後，動詞「根本」的近義詞排名第 25 名為非謂形容詞「基本」，名詞「基本」則排名落在 3200 名之後，兩者相似值僅 0.14333088954344522，可見詞性訊息確有其作用。

5. 雙語對譯：以近義詞的英文再次對應中文，以提高近義詞的召回率。

（二）測試資料及驗證結果

選取另一部分近義詞作為測試資料，用以驗證及評估以上所建立的近義詞偵測及判別原則是否適切，以下以同時出現在六部辭典的「核心」和「中心」、同時出現在五部辭典的「活潑」和「活躍」為例說明。

以「核心」和「中心」為例：第一步驟，增加詞性原則，「中心」歧義（Na 及 Nc）產生辨識性，「中心」(Na) 在「核心」的近義詞候選名單排名第 40 名；第二步驟，考慮詞性異同，剔除詞性相異者，「中心」成為第 25 名；第三步驟，以英文翻譯找到「核心」的近義詞如下：「主軸」、「關鍵」、「關鍵性」、「重心」、「中心」，「中心」排名提升至第 5 名；第四步驟，考慮詞素數目，「中心」排名升至第 4 名；第五步驟，利用雙語對譯找到更多近義詞候選詞條，包括：「重點」、「要素」。本文找到「核心」的近義詞有：「主軸」、「關鍵」、「重心」、「中心」、「重點」、「要素」，已涵蓋五部辭典所列的「中心」近義詞，亦涵蓋《教育部重編國語辭典修訂本》近義詞 [13]；此外，「重心」和「重點」則 見於《同義詞詞林》。

以「活潑」和「活躍」為例：第一步驟，增加詞性原則，「活躍」在「活潑」的近義詞候選名單排名第 1,137 名；第二步驟，考慮詞性異同，剔除詞性相異者，「活躍」排名提升至第 776 名；第三步驟，以英文翻譯找到的近義詞如下：「生動」、「好動」、「鮮活」、「爽朗」、「傳神」、「有勁」、「神氣」、「旺盛」、「鮮明」、「明朗」、「活躍」，「活躍」成為第 11 名；第四步驟，考慮詞素數目，「活躍」仍為第 11 名；第五步驟，利用雙語對譯找到更多近義詞候選詞條，包括：「積極」、「主動」、「踴躍」、「靈活」、「活絡」、「燦然」、「外向」、「開朗」、「豪爽」。本文找到「活潑」的近義詞有：「生動」、「好動」、「鮮活」、「爽朗」、「傳神」、「有勁」、「神氣」、「旺盛」、「鮮明」、「明朗」、「活躍」、「積極」、「主動」、「踴躍」、「靈活」、「活絡」、「燦然」、「外向」、「開朗」、「豪爽」，已涵蓋五部辭典所列的「活潑」近義詞，亦涵蓋《教育部重編國語辭典修訂本》近義詞（「靈活」、「活躍」、「生動」）。

13 《教育部重編國語辭典修訂本》採用「相似詞」之說法，其概念等同近義詞。

五、結論

經由本文所建立的近義詞偵測及判別規則，可搜尋到坊間辭典所提供的大部分近義詞，亦可提供更多的近義詞候選詞條。本文貢獻如下：（一）以自然語言處理和語言學分析方式處理近義詞，綜合語料庫及電子資源方法，充分反映詞彙語意、句法和語用特點，避免傳統人工編纂費時耗力及資料雜訊過多的缺點；（二）本文所建立的近義詞偵測及判別規則，經由詞性訊息、詞性異同、英文翻譯、詞素數目、雙語對譯等規則，本文提供遠較傳統方式更正確豐富的近義詞候選詞條；（三）本文所提供的近義詞可作為近義詞辨析、辭典編纂、詞彙教學及文章寫作等方面應用；（四）本文所建立的近義詞判別規則，未來可應用至發展中文近義詞自動辨識技術及系統

本研究在研究資料及方法上，受限於以下三點：（一）「華語文語料庫」自動分詞及標記之正確性若能提升，其龐大且經合法授權之語料可俾利中文自然語言處理；（二）雙語語料庫待擴充，華英對譯精確度需更細緻調校，將有助於改善目前出現無資料可搜尋或找不到對譯之情形；（三）廣義知網的部分同義詞/近義詞和坊間辭典編列的近義詞差距較大，建議廣義知網可再行編纂調整各詞項之同義詞組/近義詞組。

未來研究可進一步處理的議題，包括：（一）利用雙語對譯所找出的中文近義詞，需達到多少交集數方不致過於發散，仍待討論；（二）在以上處理訓練資料及測試資料的過程，目前人工判定及對應較為耗時，故暫設定主詞項最後一個近義詞項為坊間辭典提供的近義詞；未來自動化偵測查詢時，近義詞查詢數的準確率及召回率需再確切實驗及評估；（三）在近義詞偵測方法上，本文曾嘗試以加減法調整 Word2Vec 向量，目前動詞的近義詞判別準確率較高，名詞仍待加強；此外，本文曾嘗試以模組樣版為本(pattern-based)分析近義詞共現的句型結構，目前發現共現情形未必搭配對等連接詞（大多數對等連接詞所連接的不是近義詞關係），近義詞間較常帶標點符號「、」；未來仍需將以上方法和模組工具結合，以便調整自動偵測結果，讓查詢更為便利。

參考文獻

[1] 周荐，《漢語詞彙研究史綱》，北京：語文出版社，1995。

[2] 周祖謨，《漢語詞彙講話》，西安：人民教育出版社，1959。

[3] 王理嘉、侯學超，〈怎樣確定同義詞〉，語言學論叢，第 5 卷，頁 232-249，1963 年。

[4] 曾志雄，〈利用同義關係進行語文教學舉例〉，跨世紀的大專語文教學，香港：中文大學出版社，頁 147-154，2001 年。

[5] S. Pinker, *The Language Instinct: How the Mind Creates Language*. London: Penguin, 1995.

[6] J. R. Taylor, *Linguistic Categorization*. Oxford: Oxford University Press, 1995.

[7] J. R. Taylor, “Near synonyms as co-extensive categories: "High" and "tall" revisited,” *Language Sciences*, vol. 25, no. 3, pp. 263-284, 2003.

[8] J. I. Saeed, *Semantics*, 3rd ed. Oxford: Wiley-Blackwell, 2009.

[9] S.-F. Chung, and K. Ahrens, “MARVS revisited: Incorporating sense distribution and mutual information into near-synonym analyses,” *Language and Linguistics*, vol. 9, no. 2, pp. 415-434, 2008.

[10] E. Agirre, and G. Rigau, “A proposal for word sense disambiguation using conceptual distance,” *Proceedings of the First International Conference on Recent Advances in NLP (1995)*, Tzigov Chark, Bulgaria, 1995, pp. 258-264.

[11] P. Bhattacharyya,and N. Unny, “Word sense disambiguation and measuring similarity of text using WordNet,” *Real World Semantic Web Applications*, Amsterdam: IOS Press, pp. 3-28, 2002.

[12] 王斌，《漢英雙語語料庫自動對齊研究》，博士論文，中國科學院研究生院計算技術研究所，1999。

[13] 劉群、李素建，〈基於《知網》的辭彙語義相似度計算〉，中文計算語言學期刊，第 7 卷，第 2 期，頁 59-76，2002 年。

[14] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis*, Jersey: Lawrence Erlbaum Associates, 2007.

[15] I. Dagan, “Contextual word similarity,” *Handbook of Natural Language Processing*, New York: Marcel Dekker, pp. 474-491, 2000.

[16] I. Dagan, L. Lee, and F. Pereira, “Similarity-based models of word cooccurrence

probabilities," *Machine Learning*, vol. 34, no. 1-3, pp. 43-69, 1999.

[17] K.-J. Chen, and J.-M. You, "A study on word similarity using context vector model,", *Computational Linguistics and Chinese Language Processing*, vol. 7, no. 2, pp. 37-58, 2002.

[18] T. Wang, and G. Hirst, "Near-synonym lexical choice in latent semantic space," *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijin, 2010, pp. 1182-1190.

[19] 陳明蕾、王學誠、柯華葳，〈中文語意空間建置及心理效度驗證：以潛在語意分析技術為基礎〉，中華心理學刊，第 51 卷，第 4 期，頁 415-435，2009。

[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *Computation and Language, arXiv preprint arXiv: 1301.3781*, 2013.

[21] A. Kutuzov, and I. Andreev, "Texts in, meaning out: Neural language models in semantic similarity task for Russian," Computational Linguistics and Intellectual Technologies. *Papers form the Annual International Conference, dialogue*, vol. 14, no. 21, pp. 143-154, 2015.

[22] 蔡美智，〈華語近義詞辨識難易度與學習策略初探〉，臺灣華語教學研究，第 1 期，頁 57-79，2010。

[23] 柯華葳、林慶隆、張俊盛、陳浩然、高照明、蔡雅薰、張郁雯、陳柏熹、張莉萍，《華語文八年計畫「建置應用語料庫及標準體系」105 年工作計畫書》，國家教育研究院，2015。