

學習者詞典是針對語言學習者尤其是提供外國人學習語言的詞典。歐美國家將語料庫分析技術應用於學習者詞典的編輯已行之有年。從語料庫中觀察詞語的詞頻、語義、用法、例句及搭配詞等，以實際的語言例證做為基礎，客觀建立語言知識。而詞典輔助編輯系統的導入除了提供便利的詞典編輯介面之外，一方面讓編輯工作和語料庫分析工具緊密結合，另一方面使詞典的編輯體例維持一致、提供流程管理及版本維護等等，以提高詞典的效率與品質。

相較於國外大量採用語料庫分析技術編輯詞典的情況，以華語為主的學習者詞典仍鮮有基於語料庫的編輯方式。本研究目的為將中文語料庫分析技術導入中文學習者詞典編輯的探討，針對詞典編輯的技術做了許多面向的考察，包括一、語料庫查詢工具、二、詞典輔助編輯工具、三、中文搭配詞抽取技術及四、中文近義詞抽取技術等等。在語料庫查詢工具及詞典輔助編輯工具上，本研究依據文獻分析了現有語料庫查詢工具及詞典輔助編輯工具的三十項重要特性與功能，並以這些特性與功能，全面測試九套語料庫查詢工具及詞典輔助編輯工具。在中文搭配詞及中文近義詞抽取技術上，則以中文語料為基礎，實作並比較了十一種演算法。

經本研究實際的測試得到下列結果：

- 一、語料庫查詢工具：英國語料庫專家 Adam Kilgarriff 所發展的 Sketch Engine 是目前功能最完整的語料庫查詢工具，具備大部份重要的語言分析技術，非常適合應用在詞典編輯上。惟該系統為商業軟體，除必需考量費用支出外，自己蒐集的語料庫也必需存放在該公司的系統上。替代性方案為 CQPWeb 系統，CQPWeb 功能較 Sketch Engine 少，但為開放原始碼的免費的系統，不但語料庫可以存放在自己的伺服器上，更重要的是可依自己的需求修改與擴充系統。
- 二、詞典輔助編輯工具：由於詞典輔助編輯系統的客製化需求極高，能取得測試的系統較少、功能也較不完整。由捷克 Michal Měchura 教授所開發的 Lexonomy 系統功能相對完整，而且極具調整擴充彈性，可適應不同詞典編輯需求。
- 三、搭配詞技術：以華語語料庫實際測試 8 種搭配詞演算法發現，Dice 公式及 LLR 公式所抽取的中文搭配詞效果最好。
- 四、近義詞技術：以華語語料庫實際測試 3 種近義詞演算法發現，由 Mikolov 所提出的 word2vec 技術的效果最好，遠勝於商業化的 Sketch Engine 系統所抽出的中文近義詞。

詞典編輯技術的未來發展建議包含：

- 一、發展華語文語料庫查詢系統：現有的語料庫分析系統都忽略了中文的特性，例如離合詞的查詢功能。未來可以使用開放原始碼的系統做為發展的基礎，例如 CQPWeb 系統，發展更好的中文語料庫查詢功能。
- 二、發展中文詞彙特性速描技術：詞彙特性速描技術是非常重要的語言分析技術，由於中文和英文

的語法特性差異甚大，**Sketch Engine** 並不適合應用在中文的詞彙分析上。

資料來源

白明弘、吳鑑城（2017）。**學習者詞典編輯之研究(I)子計畫二：英語與華語學習者詞典編纂現況分析與詞典編纂技術探討**。國家教育研究院整合研究計畫案成果報告（NAER-106-24-F-1-02-02-1-05）。新北市：國家教育研究院。