

## 香港粵語前置副詞及句末成份「同句共現統計」資料集

劉擇明（香港大學中文學院漢語中心）

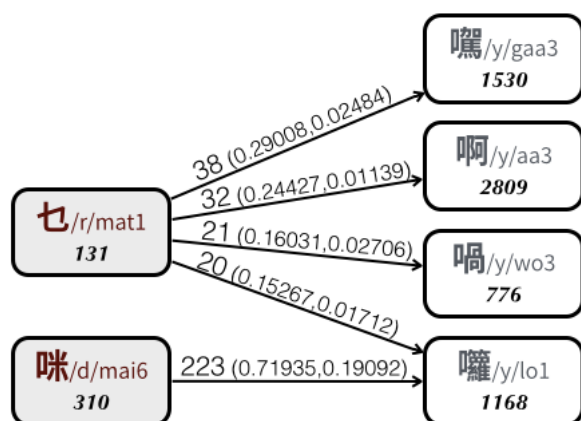
粵語中某些前置副詞和句末助詞（包括後置副詞）經常同時出現，形成「框式結構」（鄧思穎 2006）。前後成份於句中結構對稱、意思重覆，如例 (1) 所示，三種用法均見於現代粵語語料，意思通常無顯著分別：

(1a) 應該有啞。 (1b) 冇啞。 (1c) 應該冇。

然則三種用法接受度、出現頻度並不一致，副詞可以有幾個可以配搭的句末助詞；句末助詞亦可以有多個可能的前置成份配對。如能列舉配搭組合並算出前後成份出現頻率，想必有助語言處理、粵語教學、詞典編寫等資料整理。

現有粵語口語資料庫多數有斷字，詞性標註，但未必可直接用於語法分析；樹狀語料庫則制作需時。若將研究範圍縮窄，專注探討後置粵語句末助詞／動詞後綴與前置副詞的分佈及其語法，則不必建立全面標記樹狀語料庫，只需標示相關元素，即可提供豐富研究材料。中文語料庫研究中，經常用「共現度」（即兩個成份同時出現的頻度，及各自獨立出現的頻度）去判斷兩個相連字是否屬於同一詞，此一方法應可用於框式結構的資料收集。

本研究以《香港粵語語料庫》(Luke and Wong, 2015) 為原始資料，首先抽取並標出句中可用於框式結構的前置成份（如副詞、否定極項、A-not-A 等）及後置成份（如句末助詞），得出配搭合共 27 萬對，再統計出現頻度，例子，相距，以建立一資料集(dataset) (2) 如下：



範例 (2) 顯示「咪 (mai6)」「囉 (lo1)」於語料中分別出現 310 次及 1168 次，但有 223 次會同時出現，此一組合雙向的共現度皆極高（「咪」後有「囉」0.71935，「囉」前有「咪」0.19092）；「乜」於語料中出現 131 次，後續有「噃 (gaa3)」機會甚高 (0.29008)，而「噃 (gaa3)」則出現 1530 次，前面接「乜」的機率不高 (0.02484)。

本文亦會探討此資料組如何輔助各方人士釐清各個句末助詞及副詞用法、語法，與及用於詞典編寫、語言教學、電腦語言處理（尤其是倒裝句處理）等方面之實際用途。

### 參考文獻

鄧思穎(2006). 粵語框式虛詞結構的句法分析. 《漢語學報》 2 16-23.

Luke, K. K. and Wong, M. L. Y. (2015). The Hong Kong Cantonese Corpus: Design and Uses. *Journal of Chinese Linguistics* (to appear).