

LIVAC AS A MONITORING CORPUS FOR
TRACKING TRENDS BEYOND LINGUISTICS

Benjamin K. Tsou*

*City University of Hong Kong, The Chinese University of Hong Kong, and
The Hong Kong University of Science and Technology*

Oi Yee Kwong

The Chinese University of Hong Kong

ABSTRACT

After almost two decades of development, the LIVAC corpus¹ has matured from a synchronous corpus to become a large diachronic corpus, capable of tracking not only linguistic variations across various Chinese speech communities but also other related societal and cultural changes across the Chinese language. This paper[†] examines how LIVAC can also function effectively as a monitoring corpus to allow both latitudinal and longitudinal analyses for various linguistic phenomena, especially lexical semantic changes as a result of cultural contact. We also revisit the important issue of minimum character and word thresholds for literacy, based on more updated empirical data offered by LIVAC and other considerations.

KEYWORDS

Monitoring corpus Latitudinal analysis Longitudinal analysis
Pan-Chinese speech communities Literacy threshold in Chinese

* Corresponding author: btsou99@gmail.com

† Section 4 of this paper (RECIPROCITY IN CULTURAL CONTACT) is in part based on Tsou and Tsoi (2013) which first discussed some of the relevant facts.

從LIVAC追蹤語料庫探索泛華語地區語言以外的演變

鄒嘉彥

鄭藹兒

香港城市大學, 香港中文大學, 香港中文大學
及香港科技大學

提要

LIVAC 自1995年起開發, 至今已從一個共時語料庫發展成龐大的歷時語料庫, 緊貼社會變遷, 將近廿年間在泛華語地區出現的語言現象, 反映於客觀的數據中, 為語言和社會發展等研究提供了重要的材料。本文以數例闡述如何以LIVAC作為追蹤語料庫, 將泛華語地區詞彙語義轉變等語言現象與社會因素結合, 進行橫向及縱向分析, 以達至全面的理解。我們亦根據最新和更大量的LIVAC數據, 分析字詞數量與閱讀能力的關係, 重新審視過去以三千常用字為掃盲標準的說法。

關鍵詞

追蹤語料庫 橫向分析 縱向分析 泛華語地區 掃盲標準