

**Second Asia Pacific Corpus Linguistics Conference  
(APCLC 2014)**

**Title: A Preliminary Study on Computer-Aided Acquisition of Cantonese Vocabulary  
by Mandarin Speakers using a Parallel Corpus**

**Theme: corpora and language learning**

Wong Tak-sum\* and John Lee†  
The Hong Kong Polytechnic University\*  
City University of Hong Kong†

**Abstract**

While there have been studies on computer-aided learning of Mandarin vocabulary (e.g. Yang & Xie 2013), less attention has been paid to Cantonese. In this paper, we report our experience in teaching Cantonese vocabulary to Mandarin-speakers with the aid of the first parallel corpus that aligns transcribed Cantonese speech and the written Mandarin counterpart (Lee 2011).

In general, a polysemous word in one language often has different counterparts in another language for each meaning. This often poses obstacles to learners. We have thus conducted an experiment by inviting a group of Mandarin-speaking undergraduate students from a Cantonese class at City University of Hong Kong to look for the Cantonese counterparts of some polysemous words in Mandarin in our corpus as well as to analyze their meanings or functions (Figure 1).

The effect of the use of corpora on our subjects' proficiency in Cantonese was measured via pre-tests, post-tests and delay post-tests. We have administered a pre-test to collect a set of Cantonese words that each student did not understand. In personalized computer-assisted vocabulary learning session, given the Mandarin counterpart, each student was asked to search for half of these words in the corpus. There exists statistically significant difference between the students' mastery of the words that were included in the computer-assisted section and the words that were not, in both post-test ( $p < 0.0005$ ) and delay post-test ( $p < 0.05$ ). It can be seen that our corpus is an effective device of computer-aided acquisition of Cantonese vocabulary.

<u>Mandarin polysemous words given</u>	<u>Cantonese counterparts to look for</u>	<u>meaning / function</u>
了 <i>le</i>	咗 <i>jó</i> 嘍 <i>la</i>	<i>perfect aspect particle</i> <i>mood particle</i>
對 <i>duì</i>	啱 <i>ngāam</i> 對 <i>deui</i>	'correct' 'to face'
會 <i>huì</i>	會 <i>wúih</i> 識 <i>sík</i>	<i>future marker</i> 'to know'

Figure 1 Examples of Mandarin polysemous words and their Cantonese counterparts

**References**

- Lee, J. 2011. Toward a parallel corpus of spoken Cantonese and written Chinese. In: Proceedings of the Fifth International Joint Conference on Natural Language Processing, Chiang Mai, Thailand, November 8 – 13, pp. 1462–1466.
- Yang, C. and Xie, Y. 2013. Learning Chinese Idioms through iPads. *Language, Learning & Technology*, 17(2), 12–22.