

## **Cantonese Final-Initial Syllable Biphone Probability in Wordlikeness Judgement**

**Tak Wang LI <sup>a</sup>, Winsy Wing Sze Wong <sup>b</sup>**

*<sup>a</sup>Incoming Ph.D. Student, Department of Speech and Hearing Sciences,  
University of Washington*

*<sup>b</sup>Research Assistant Professor, Faculty of Humanities, Department of Chinese and Bilingual  
Studies, The Hong Kong Polytechnic University*

### **Abstract**

This study investigates the role of cross-syllabic transitional probability in lexical acceptability. Lexical judgement is theorised in a Boolean and deterministic manner by classical generative phonology, either grammatical or ungrammatical. Yet, probabilistic phonology has forwarded the gradient or probabilistic nature of linguistic presentations as they are regulated by statistics and probability, with a spectrum of goodness. The role of probability has been extensively studied in precursor literature showing certain effects such as higher acceptability ratings (e.g. Bailey & Hahn, 2001). Such gradient knowledge is commonly captured with the wordlikeness rating paradigm where participants rate the similarity of an item to a specific language.

Previous research on Cantonese probabilistic representations has primarily focused on lexical gaps (Kirby & Yu, 2007) and tones (Do & Lai, 2020). Nevertheless, cross-syllabic transitional units or biphones were never considered. Biphone probability matters much in speech synthesis development (Law & Lee, 2000), speech boundary detection (Chambers et al., 2002), and potentially in studying phonological disorders in individuals with aphasia and their rehabilitation (e.g., Kendall & Nadeau, 2016). Given the abundance of bisyllabic words in Cantonese, it is essential to understand how mental lexical representations are organised based on biphone probability.

Addressing this gap, a wordlikeness rating task was conducted with 20 adult native Cantonese speakers. In the preliminary stage, text data were amassed from the audios of City Forum, a television show where public figures discuss social issues in Cantonese. This show hence provides authentic representations of spoken Cantonese. The data were tokenised into Jyutping in ELAN. Biphone patterns were then extracted and calculated from the dataset in Sublime Text and with R code. The probability of each biphone was eventually calculated by dividing the sum of the frequency of a specific sound sequence by the total frequencies of all items (Faizal et al., 2015).

300 stimuli were created and were meaningless disyllabic alongside evenly distributed biphone probability. Two experimental conditions are created considering the effects of lexical gaps in previous research (Kirby & Yu, 2007). The first contains stimuli created with existing syllables; the second contains stimuli created with non-existing syllables, be them systematic and accidental gaps. During the experiment, participants evaluate the Cantonese-likeness of the stimuli on a scale of 1 to 7.

Our mixed-effect regression model shows that biphone probability positively and significantly correlates to wordlikeness rating and lexicality is a discriminating factor in wordlikeness rating: rating in the existing condition is significantly higher than in the non-existing condition. Our post-hoc analysis reveals that the significance between biphone probability and rating is only present in the existing syllable condition. This implies that biphone probability may affect lexical processing of bisyllabic words, which supports the multi-level cluster model of Chinese mental lexicon proposed by Zhou and Marslen-Wilson (1995). The non-existing condition presents a positive yet insignificant correlation. This insinuates much weaker lexical-phonological activations presumably stem from the nature of the stimuli and possibly a small sample size.