

Tonal aspects of code-switching: Three case studies of English-Cantonese/Mandarin/Vietnamese

Katrina Kechun Li¹, Christopher Bryant², Kayeon Yoo³, Li Nguyen²¹TAL ²ALTA, DCST ³Amazon

Tonal effects in Code-switching

Code Switching:

- 你會即係 closer to another circle 囉 [Cantonese-English] (HLVC corpus)
'You will then be closer to another circle (sentence-final particle).

Phonological effect in code-switching:

- Olson (2013): Code switching influences speech production, e.g. voice onset time (VOT) in the L1 shift in the direction of the L2 norms in Spanish-English code switching.

Tonal effect in code-switching:

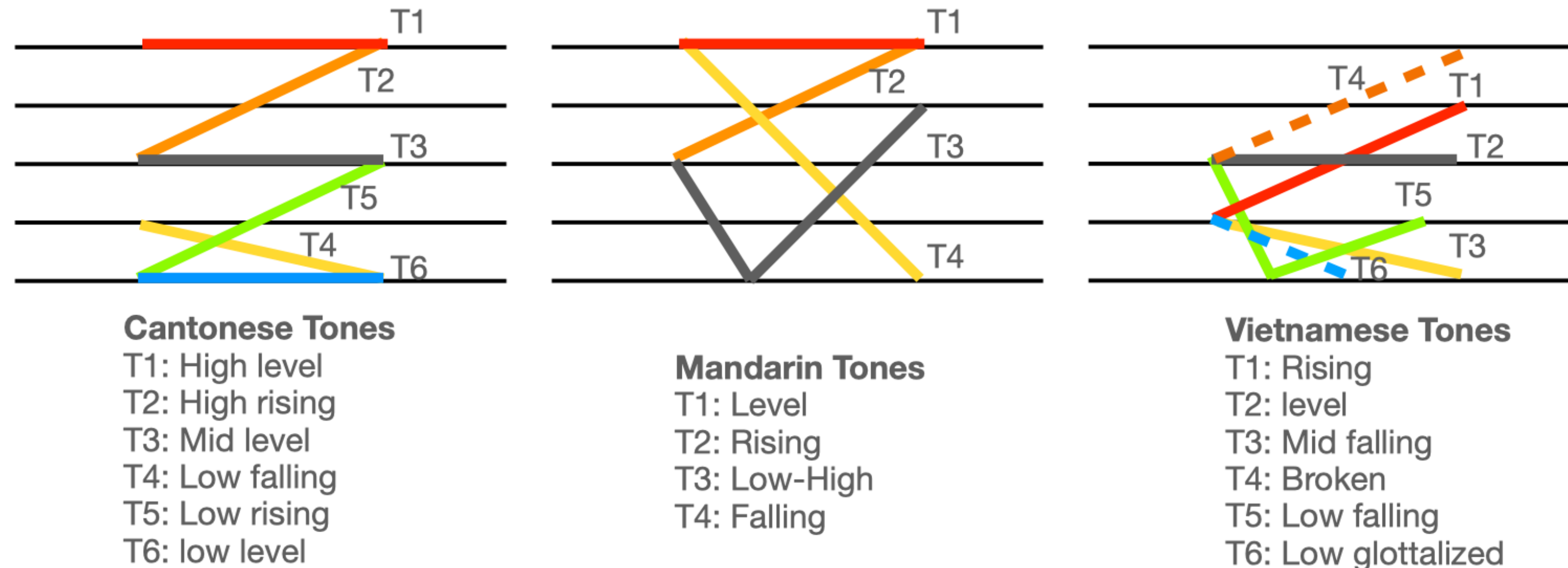
- Tuc (1997): Vietnamese into English switches tend to follow mid- to high-level pitch tones
- Zheng (1998): Chinese to English switches usually occurs following the Chinese fourth, half third and neutral falling tones or following the weak stress.

These studies, however, were small in scale and did not consider other confounding factors such as tone frequencies and syntactic structure.

Corpora

Data: Three corpora, collected from bilingual communities.

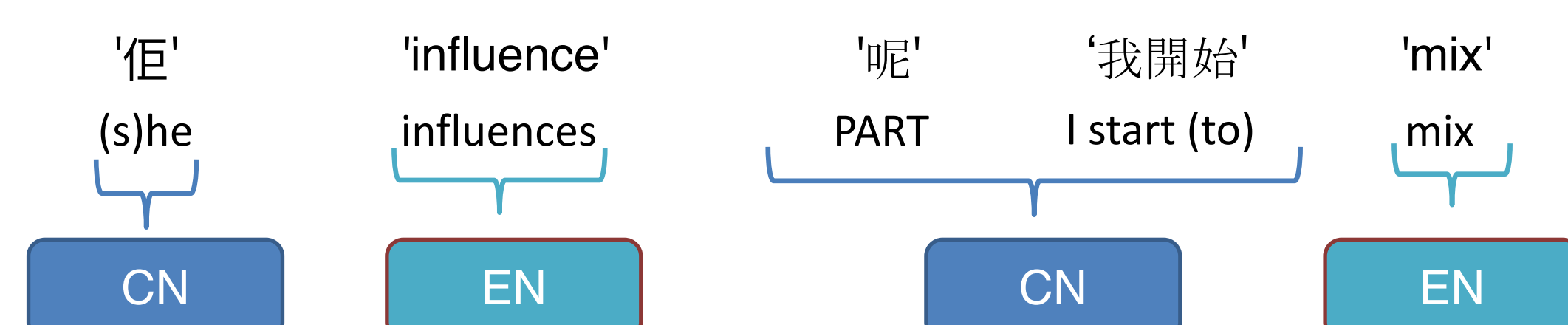
- Vietnamese-English: CanVEC (Nguyen and Bryant 2020). 45 speakers.
- Cantonese-English: HLVC (Nagy 2011). 25 speakers.
- Mandarin-English: SEAME (Lyu et al. 2010). 20 speakers.



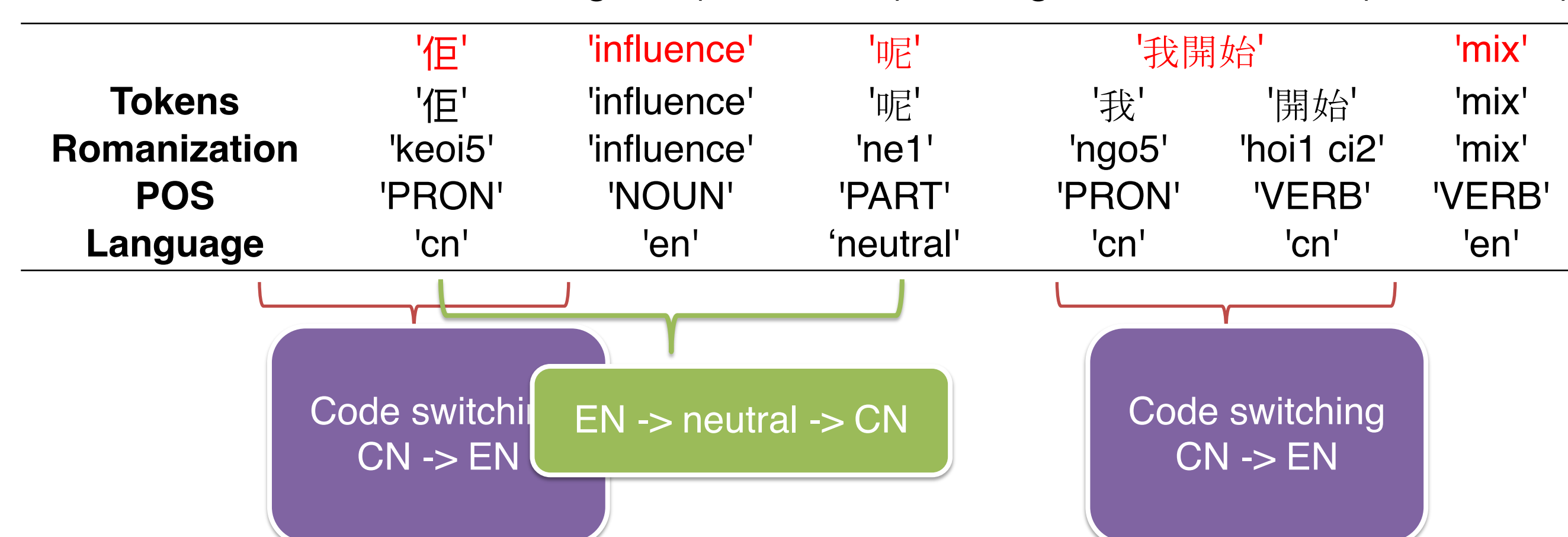
Processing

Semi-automatic natural language processing method developed by Nguyen and Bryant (2020) for Vietnamese-English.

- Based on the character encodings (English letters/Chinese characters), we divided the largest contiguous sequences of Chinese or English text in a sentence.



- The sequences of different languages were tokenised, part-of-speech tagged and romanised with PyCantonese (Lee 2015) for Cantonese, SpaCy (honnibal and Montani 2017) for Mandarin, vPhon (Kirby 2008) for Vietnamese.
- We assigned language label to tokenized data:
English, Cantonese/Mandarin/Vietnamese, **Language-neutral***
* Language neutral terms: filler words (OK, Yeah), sentence-final particles (lah, ba), proper nouns (Hong Kong, Disney).
- Code-switching points were identified based on language labels, and were marked as Chinese to English (CN -> EN) or English to Chinese (EN -> CN).

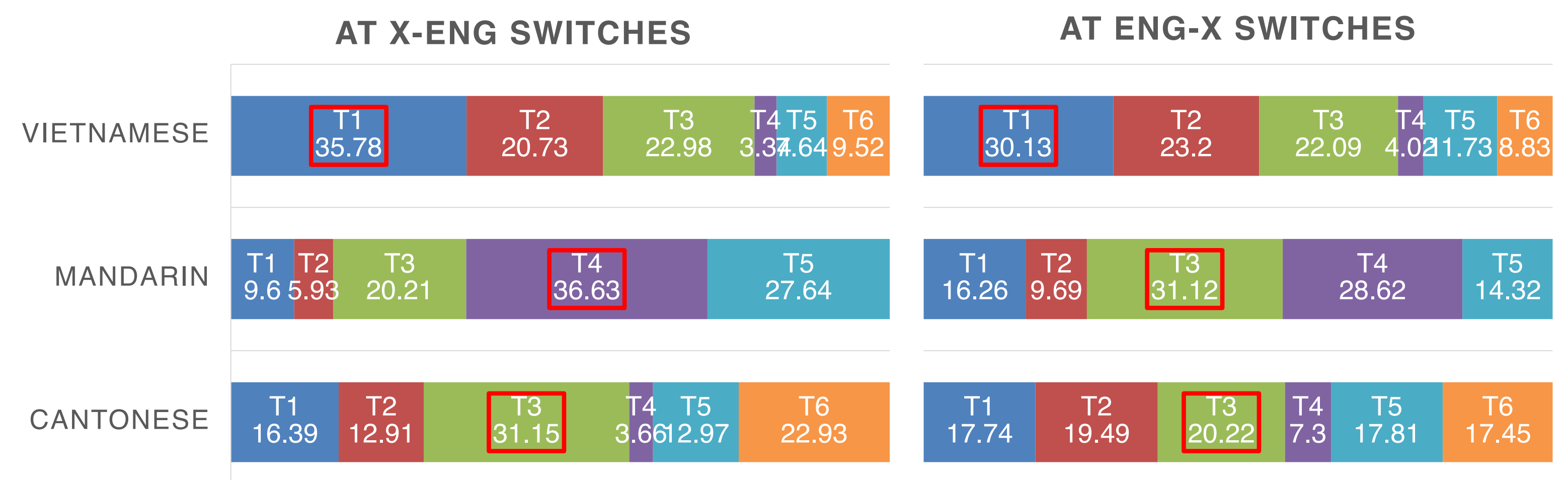


English translation: (Under) his influences, then I start to mix.

Results

Tonal Distribution at switch points

- Distribution is similar to that reported in Tuc (1997) for Vietnamese and Zheng (1998) for Mandarin



Tone occurrences at switch points relative to overall occurrences

	Vietnamese-ENG		Cantonese-ENG		Mandarin-ENG	
	Count	%	Count	%	Count	%
T1	1082	6.73	273	0.92	691	5.41
T2	627	3.84	215	0.96	427	4.06
T3	695	4.9	519	2.36	1455	6.82
T4	101	4.67	61	0.43	2638	9.29
T5	231	4.95	216	1.1	1990	17.29
T6	288	4.64	382	1.24		

	ENG-Vietnamese		ENG-Cantonese		ENG-Mandarin	
	Count	%	Count	%	Count	%
T1	809	5.03	243	0.82	1222	9.57
T2	623	3.81	267	1.19	728	6.92
T3	593	4.18	277	1.26	2339	10.97
T4	108	4.99	100	0.7	2151	7.58
T5	315	6.76	244	1.25	1076	9.35
T6	237	3.82	239	0.77		

Statistical Analysis: Logistic mixed effects modelling

- Dependent variable: whether a token is a switch point to/from English
- Fixed effects: Tone, word frequency, POS, speaker gender and generation
- Random effects: (1 + Tone|Speaker) + (1|Token)

Post-hoc pairwise comparisons:

- Vietnamese-English/English-Vietnamese: None
- Mandarin-English: T4 > T1, T3, T5; English-Mandarin: T4, T5 > T2
⇒ A facilitating effect of T4, an inhibiting effect of T2
- Cantonese-English: T3 > T1, T2, T4, T5, T6; T3, T5, T6 > T4;
English-Cantonese: T2 > T4, T5,
⇒ A facilitating effect of T3 (and marginally T2), an inhibiting effect of T4

Discussion and Conclusion

Explanation for tonal patterns

- The previously observed tonal effects in Vietnamese are side effects of some highly frequent tokens.
- The facilitatory effect of T4 (falling tones) in Mandarin is due to its similarities to the English intonation contour.
- The facilitatory effect of T3 (level tones) in Cantonese resonates with word prosody of Cantonese English words which assign level tones to English syllables (Gussenhoven 2012).
- Triggering hypothesis (Clyne 2013): Certain lexical items can act as triggers for CS in bilingual speech because of the similarity in the surface form of a trigger word across languages.

Tonal effects in code-switching is language-dependent.