

Variation of the third-person singular pronoun in Hong Kong Cantonese

Litong Chen, *The Ohio State University*

Background: The third-person singular pronoun 佢 in Hong Kong Cantonese has several variations, including the more standard (citation) form *koei* and its variant forms, such as *hoei*, *toei*, and *oei*. Some speakers use both *koei* and other forms alternatively, while some others consistently use only one of them (Bauer and Benedict 1997). This variation phenomenon is widely observed in current Hong Kong and has been discussed in a few studies. Bauer and Benedict (1997) found that younger speakers tend to use the *hoei* form more often than the older generation does. Bourgerie (1990) discovered a correlation between the use of *hoei* and both age and formality of context. He suggested that the use of *hoei* is more frequently observed among the younger generation and in more informal situations than otherwise.

Current study: Based on the findings of previous studies, this paper aims to examine other social and linguistic factors, besides age and formality of context, that influence Hong Kong Cantonese speakers' choice between the citation form *koei* (henceforth the *k*-form) and other variant forms. In particular, this paper investigates whether factors such as age, gender, roles in conversation, and/or the position of the third-person singular pronoun in sentence are closely associated with the use of the citation form vis-à-vis the variant forms.

Methods: This study collected data from the Hong Kong Cantonese Adult Language Corpus (HKCAC). This database recorded some spontaneous phone-in programs and forums on the radio of Hong Kong between November 1998 and February 2000. In total, 721 tokens of the third-person singular pronoun were extracted from the HKCAC database. This study examined two of the phone-in radio programs (all together 142 minutes long). Content details of these programs are displayed in Table 1 (c.f. Leung and Law 2001).

All tokens were then put into a dataset, which shows information of each token and its speaker. The response variable is the third-person singular pronoun (*k*-form vs. non-*k*-forms). There are four fixed variables, namely, "host/caller status", "age", "gender", and "position in sentence." There is also a random variable: "speaker." Levels of each variable are listed in Table 2. I first cross-tabulated the dataset and checked any collinearity between each pair of the independent variables. With redundant predictors removed, next, I included random variables ("speaker") and conducted a mixed-effects regression model for comparison, given that repeated measurements were made on the same speakers. Finally I interpreted the results and discussed some related findings.

Results: Cross tabulation indicates no collinearity between most variables. The only exception is "host/caller status" versus "gender." With female hosts being absent from the data, "host/caller status" becomes partly predictable by "gender." Cross tabulation suggests a removal of "gender" from independent variables, and this is confirmed by stepwise selection, as shown in Table 3. Such removal also echoes Bourgerie's (1990) study which shows no significant difference between male and female speakers in favor of one form over another. The other three predictors ("host/caller status", "age", and "position in sentence") are kept to build a mixed-effect model together with "speaker" as a random variable.

The results show that in the mixed-effects model, *k*-form is not significantly correlated to "host/caller status" or "age", as suggested by large *p*-values (0.147 and 0.106, respectively). However, *k*-form is significantly correlated to "position in sentence" ($p < 0.001$), even after considering the random effect of "speaker." As shown in Figure 4, *k*-form can be predicted by the independent variable "position in sentence" in a way that when the third-person singular variable is in a sentence-initial position, all speakers, with only two exceptions (LX and YS), tend to use the *k*-form. However, this is not the case in non-sentence-initial positions. First, a third of the speakers (10 out of 30) prefer using non-*k*-forms if not in the sentence-initial position. Second, distribution of the *k*-form and non-*k*-forms frequency between 0 (not using the *k*-form at all) and 100% (always using the *k*-form) is overall even. Both observations indicate a lower predictability of the third-person singular pronoun forms in non-sentence-initial positions.

Conclusion: This paper studies the Hong Kong Cantonese third-person singular pronoun and investigates some of the potential social and linguistic factors that are correlated with the occurrence of the *k*- vs. non-*k*-forms. The results show no significant correlation between any of the social factors involved in this study and the variable, and this finding does not correspond to results in previous studies. The only clear correlation is found between the variable and the position in sentence where it occurs, with "position in sentence" *per se* being a linguistic predictor. This paper hence suggests that the use of the citation form vis-à-vis variant forms of the Hong Kong Cantonese third-person singular pronoun is more likely to be a linguistic issue than a typical sociolinguistic problem. The fewer occurrences of non-*k*-forms in sentence-initial positions might be subject to the universal phrase- or utterance-initial blocking of lenition (Kirchner 1998).

Program	Theme	# of callers		Callers' age	Hosts		Total time
		Female	Male		Female	Male	
<i>To appease your mind</i>	Current affairs	6	9	Mid-aged	0	1	69 minutes
<i>Star trek and Titanic</i>	Personal matters	11	4	Teenagers & young adults	0	1	73 minutes

Table 1. Details of the examined programs

Independent variable	Level	
Speaker	"AE" "AG" ... "ZS" "ZX"	30
Host/caller status	Host, Caller	2
Age	Teenagers & young adults, Mid-aged individuals	2
Gender	Male, Female	2
Position in sentence	Sentence-initial, Non-sentence-initial	2

Table 2. Levels of each independent variable

K	~	host/caller	+	gender	+	age	+	position	(AIC)
P-value		<0.001		0.778		0.004		0.024	-1135.17
K	~	host/caller	+			age	+	position	(AIC)
P-value		<0.001				0.001		0.025	-1137.10

Table 3. Model comparison with and without "gender" as a predictor

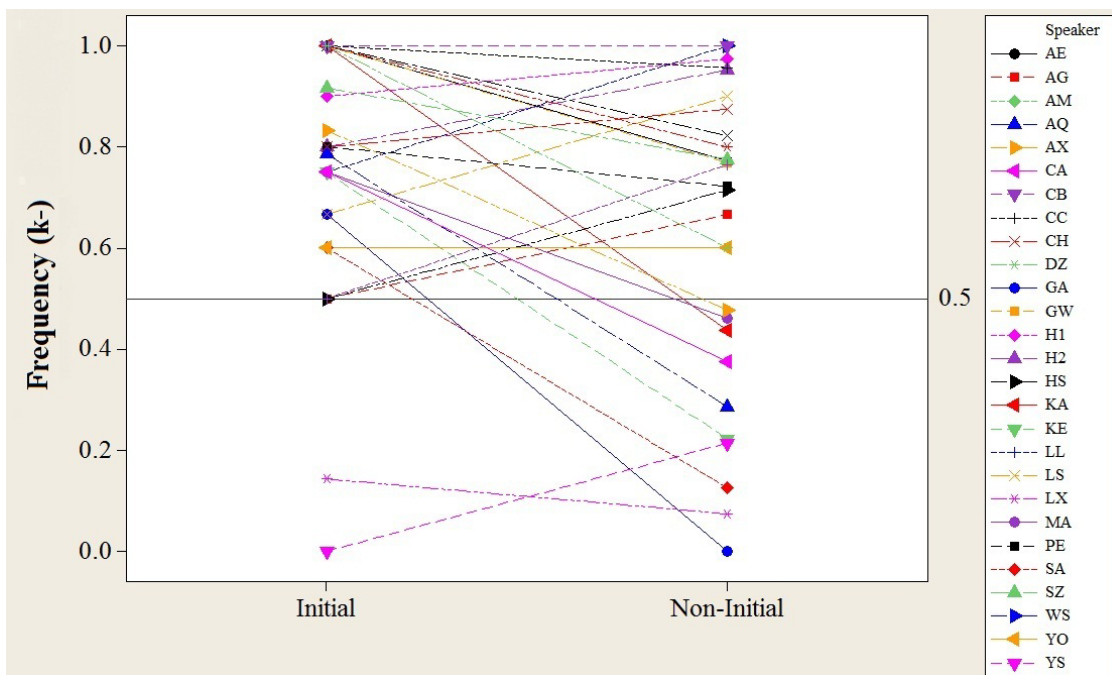


Figure 4. Line plot of frequency of koi in different positions of sentence

Selected references

- Bauer, Robert S. and Paul K. Benedict. 1997. *Modern Cantonese Phonology*. New York: Mouton de Gruyter.
- Bourgerie, Dana S. 1990. *A Quantitative Study of Sociolinguistic Variation in Cantonese*. Doctoral dissertation. The Ohio State University.
- Kirchner, Robert M. 1998. *An Effort-Based Approach to Consonant Lenition*. PhD dissertation. UCLA.
- Leung, M.-T. and S.-P. Law. 2001. HKCAC: The Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics* 6, 305-25.