Probability matching vs. regularization in contact-induced syntactic change

Ming Xiang[1], Christine Gu[1], Yixue Quan[2], Weijie Xu[1], Suiping Wang[2]
[1]The University of Chicago; [2]South China Normal University

In statistical learning, both probability matching and (over-)regularization have been found in human behavior [1]. In the former, humans reproduce the probability distribution in the input; whereas in the latter, the frequent pattern in the input is produced even more frequently than its input frequency. Statistical learning has been suggested as a possible mechanism for language change, but the conclusions are often based on artificial language learning tasks. The current study looks at syntactic change due to language contact in multilingual communities. Using a picture-description production task, we investigated the usage of ditransitive verbs across multiple generations of Cantonese speakers from Guangzhou, China. Cantonese is the major local language spoken in Guangzhou, but its usage is in decline amid intensive contact with Mandarin Chinese. The current study, being one of the first to quantitatively evaluate syntactic change in Cantonese, revealed that the younger generation of Cantonese speakers, instead of shifting to a direction that probability-matches the distribution of Mandarin (the dominant contact language), actually over-regularized the originally preferred pattern within Cantonese.

Procedure: Two main groups of participants were tested on the same set of stimuli (Table 1). The **target Group 1** are native Cantonese speakers (18-70 years old) that were born and raised in Guangzhou and currently live there. The second **control Group 2** are native Beijing Mandarin speakers (18-60 years old) that were born and raised in Beijing and currently live there. In a picture description task, participants used a verb provided to them in their respective native language to describe a picture that depicts a ditransitive event. The critical trials (n=21) all have verbs that can be used ditransitively. There are an additional 20 filler trials.

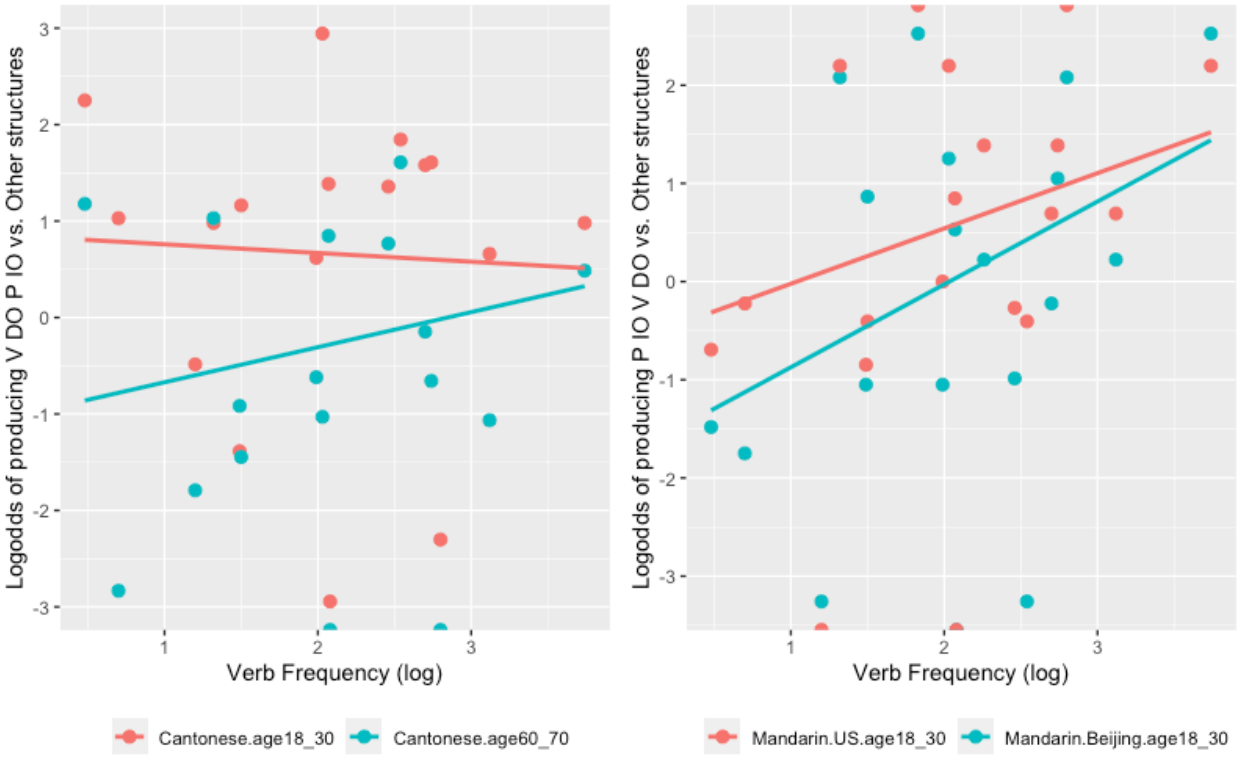
Results: Although the range of possible syntactic frames produced by the two groups of speakers are largely identical, there is a sharp contrast in the distribution of the patterns. The most frequently produced Cantonese structure is V DO P IO (55% on average, *sent some apples to the friends*), whereas for Mandarin it is P IO V DO (46%, *to the friends sent some apples*). As shown in Table 1, for Cantonese, the production frequency of the dominant V DO P IO order gradually increased from older to younger generations, and there is no change in the P IO V DO frequency, showing no assimilation to Mandarin. A mixed-effects logistic regression model, with age as a continuous variable, confirmed that older adults produced fewer V DO P IO structures (Est=-0.027, SE=0.005, z=-5.5, p<.0001). Younger Cantonese speakers therefore have **over-regularized** the originally preferred V DO P IO pattern. For each generation of Cantonese speakers, we also calculated an entropy measure based on the frequency (aggregated by participants and items) of each syntactic frame produced. We found entropy reduction from older to younger generations (Table 1), consistent with the observation of over-regularization. When parallel analyses were carried out for the group of Beijing Mandarin speakers, there is no evidence for any change at all. The over-regularization in Cantonese is therefore not the result of a universal diachronic process. It appears to have taken place because Cantonese is under the pressure of being in contact with another dominant language. To explore whether contact-induced over-regularization is a more general phenomenon, we conducted a pilot study on a third group of young Beijing Mandarin speakers, between 18-30 years old, who were born and raised in Beijing but moved to Chicago in their late adolescence or early adulthood. Compared to the age-matched Mandarin speakers living in Beijing (18-30 years), the Mandarin speakers in the US showed clear over-regularization, producing significantly more instances of the P IO V DO pattern (Est=0.75, SE=0.34, z= 2.1, p<.05). Regularization has been argued to be frequency-dependent in some previous studies [2,3]. To understand the individual item effect, we correlated each verb's **lexical frequency** with its log-odds of being used in the most dominant syntactic frame (Figure 1). We did not find an interaction between lexical frequency and old/young Cantonese groups (p>.4), nor did we find an interaction between lexical frequency and US/Beijing Mandarin groups (p>.7).

Conclusion: In an intensive multilingual environment, the weaker language does not necessarily assimilate to the dominant language. Instead we observe contact-induced (over-)regularization, suggesting a potential relationship between regularization and cognitive load pressure [4].

Table 1:

		Mandarin speakers				Cantonese speakers from Guangzhou			
	Age group	# of participants	Frequency of V DO P IO	**Frequency of P IO V DO**	Syntactic frame entropy	# of participants	**Frequency of V DO P IO**	Frequency of P IO V DO	Syntactic frame entropy
	60-70	NA				20	**0.4**	0.06	2.68
Beijing	50-60	14	0.04	**0.45**	2.48	23	**0.58**	0.07	2.23
	40-50	17	0.05	**0.47**	2.45	23	**0.62**	0.03	2.07
	30-40	22	0.11	**0.47**	2.42	16	**0.63**	0.02	1.98
	18-30	27	0.05	**0.47**	2.52	22	**0.67**	0.08	1.94
US	18-30	10	0.03	**0.57**	2.12	NA			

Figure 1: Relationship between lexical verb frequency and the log-odds of producing the dominant syntactic frame over other structures. **Cantonese (Left)**: oldest and youngest generations. **Mandarin (Right)**: speakers living in the US and their age-matched counterparts living in Beijing.



References:
[1] Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development,*1(2), 151-195.
[2] Morgan, E., & Levy, R. (2016). Frequency-Dependent Regularization in Iterated Learning. In *The Evolution of Language: Proceedings of the 11th International Conference*.
[3] Liu, Zoey and Morgan, Emily (2020) Frequency-dependent Regularization in Constituent Ordering Preferences. *The 42nd Annual Conference of the Cognitive Science Society*.
[4] Ferdinand, V., Kirby, S. & Smith, K. (2019) The cognitive roots of regularization in language. *Cognition*.