USING COMPUTER SIMULATION TO STUDY LINGUISTIC DIFFUSION

Tao Gong   Lan Shuai
Haskins Laboratories, New Haven, USA

ABSTRACT

Human language is constantly changing. Many of the changes are achieved via diffusion of linguistic variants in a population of interacting individuals. We design a simple computer model following the Pólya-urn dynamics to evaluate possible effects of linguistic, individual learning, and socio-cultural factors on linguistic diffusion. Simulation results and statistical analyses collectively reveal that: variant prestige, rather than individual bias, is a fundamental selective pressure for diffusion; given variant prestige, speaker's and hearer's preferences for prestigious variants show different degrees of efficiency on diffusion; and structural features of social network, such as average degree and level of centrality, also modulate the degree of diffusion. This work offers an interdisciplinary approach to study linguistic diffusion and modify

Gong, Tao 龚涛 is a research scientist at Haskins Laboratories, Yale University, New Haven, CT; [gong@haskins.yale.edu]. He is the author of *Computational Simulation in Evolutionary Linguistics: A Study On Language Emergence* (2009) and numerous interdisciplinary articles on computational simulation of language learning and evolution in journals like *Proceedings of the National Academy of Sciences of the USA*, P*roceedings of the Royal Society B: Biological Sciences, Physics of Life Reviews, Language, Lingua,* or *PLOS ONE*. His recent research focuses on connectionist models and psychological experimental studies of reading development and individual difference.

Shuai, Lan 帅兰 is an affiliated scientist at Haskins Laboratories, Yale University, New Haven, CT; [shuai@haskins.yale.edu]. She is the author of numerous interdisciplinary articles on neuroimaging studies of lexical tone processing and simulation studies of language learning and evolution in journals like *Proceedings of the Royal Society B: Biological Sciences, Frontiers in Behavioral Neuroscience, Frontiers in Neuroscience, PLOS ONE*, or *Physics of Life Reviews*. Her recent research focuses on neuroimaging studies of reading development and language disorders.

available theories and simulations concerning individual learning or social factors.

KEYWORDS
Linguistic diffusion **C**omputational modeling **P**ólya-urn dynamics **C**omplex networks

1. EVOLUTIONARY LINGUISTICS AND COMPUTATIONAL MODELING

Language is arguably the very cognitive capacity that defines humans (Beckner et al. 2009). As a socio-cultural phenomenon (Hruschka et al. 2009), language undergoes constant changes, many of which proceed via diffusion of phonetic, lexical, or syntactic variants, such as *the Great Vowel Shift* in English in the 14-16th century (Wolfe 1972), sound change in *Wu* dialect of Chinese (Shen 1997), or lexical borrowing (Cheng 1987). At the population level, *linguistic diffusion* (henceforth "diffusion") can be defined as a shift in proportions of linguistic variants used by a population over time (Nakamura et al. 2007). Studying the manners in which diffusion takes place can yield important insights on questions concerning the cognitive capacities for language and linguistic or socio-cultural constraints on language evolution (Pinker and Bloom 1990; Croft 2000; Labov 2001; Hauser et al. 2002; Tomasello 2008).

Many diffusion theories assume *an innovator-learner framework* to discuss who innovates or introduces preference for certain types of variants, who learns variants during linguistic interactions (Mufwene 2008), and what factors influence individuals' choices among variants (Nettle 1999; Labov 2001). Many empirical studies in sociolinguistics (Fisiak 1995; Croft 2000; Labov 2001) have revealed that linguistic, individual-learning, and socio-cultural factors all affect diffusion. Nonetheless, there remain some questions, including:

(1) What is the *selective pressure for diffusion* (factor(s) leading to successful spread of variants in a population), variant prestige or individual bias?

(2) Does the speaker's or hearer's preference for certain types of variants show different effects on diffusion?

Nature, ed. M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. M. Guervós, and H-P. Schwefel, 17-26. Berlin: Springer-Verlag.

TOMASELLO, Michael. 2008. *The Origins of Human Communication*. New York: MIT Press.

WOLFE, Patricia M. 1972. *Linguistic Chang and the Great Vowel Shift in English*. Berkeley, CA: University of California Press.

ZHANG, Menghan and Tao Gong. 2013. Principles of parametric estimation in modeling language competition. *Proceedings of the National Academy of Sciences of the United States of America* 110(24):9698-9703.

用模拟仿真研究语言扩散

龚涛　　帅兰

美国 Haskins 实验室

摘要

人类语言不断变化。很多变化是通过在一组交流个体中扩散不同形式的语言变体完成的。我们设计了一个基于 Pólya-urn 动态的简单计算机模型研究语言、个体学习和社会-文化因素对语言扩散可能起的作用。模拟结果和统计分析表明：（1）对变体的偏好，而不是对使用者的偏好，是决定扩散的基本选择要素；在存在变体偏好情况下，（2）说者和听者对变体的偏好具有不同程度的扩散效率；（3）网络结构因素，如平均度和集中度，可调节扩散程度。此工作为研究语言扩散提供了一个跨学科手段，并改进了已有的关于个体和社会因素对语言扩散影响的理论和模拟研究。

关键词
语言扩散　　**计**算机模拟　　**P**ólya-urn 动态　　**复**杂网络