INDIVIDUAL AND SOCIAL EFFECTS ON LINGUISTIC DIFFUSION

TAO GONG
Department of Linguistics, University of Hong Kong, Pokfulam Road, Hong Kong

MÓNICA TAMARIZ
Department of Linguistics and English Language, University of Edinburgh Charles Street, Edinburgh, UK

GERHARD JÄGER
Department of Linguistics, University of Tübingen, Wilhelmstraße, Tübingen, Germany

Using a Pólya urn based diffusion model, we examine linguistic diffusion in a population of individuals with different types of social networks. Simulation results and statistical analyses show that individual factors, such as who (speakers or hearers) introduce preference for certain types of variants, and network structural features, e.g., level of centrality, collectively modulate the degree of diffusion. This work shows different diffusion efficiencies of speaker's and hearer's preferences for variants, evaluates the effect of individual learning and social factors on linguistic diffusion, and modifies previous diffusion theories focusing exclusively on individual or social factors.

1. Introduction

Language is a socio-cultural phenomenon that undergoes constant changing (Hruschka et al., 2009). Many changes are achieved via diffusion of various types (phonetic, lexical, syntactic, etc.) of linguistic variants, such as the Great Vowel Shift in English in the 14-16th century (Wolfe, 1972), other sound change in *Wu* dialect of Chinese (Shen, 1997), or lexical borrowing (Cheng, 1987). *Linguistic diffusion* (henceforth "diffusion") can be viewed as the shift in proportions of linguistic variants used by a population over time (Nakamura et al., 2007). Studying the manners in which diffusion occurs helps understand human cognitive capacities for language and socio-cultural constraints on language evolution (Croft, 2000; Labov, 2001).

Theories of diffusion usually assume an innovator-learner framework, discussing who mainly innovate or introduce preference for certain types of

variants and who mainly learn variants in linguistic interactions (Mufwene, 2008) or examining factors that can affect individuals' choices of variants or languages (Labov, 2001; Nettle, 1999). Besides the small-scale, empirical approach in sociolinguistics (Croft, 2000; Fisiak, 1995; Labov, 2001), mathematical and behavioral simulations have been recently adopted to study the effect of linguistic, individual learning, and socio-cultural factors on language evolution. By quantifying contact patterns and social prestige, mathematical models help predict the outcome of language competition (e.g., Abram & Strogatz, 2003; Minett & Wang, 2008); by simulating linguistic behaviors, behavioral models trace the conventionalization of form-meaning mappings (e.g., Steels, 1995) and linguistic regularities (e.g., Gong, 2011) via local interactions, and reveal the correlation between population size or social connections and individual memory load or diffusion success (e.g., Dall'Asta et al., 2006; Ke et al., 2008). However, these studies usually separate individual and social factors in diffusion. For example, in mathematical models, individual choice of languages is mainly determined by global factors such as linguistic prestige, disregarding social connections. In behavioral models, hearers mainly acquire and update linguistic knowledge. Studies that connect networks with learning factors (e.g., Dall'Asta et al., 2006) only examine simple networks such as lattice networks, and others that consider complex networks (e.g., Ke et al., 2008) lack a systematic analysis of the effect of structural features on diffusion.

During diffusion, speakers or hearers, or both, can innovate or introduce identical or different preferences for variants. For example, speakers tend to use easily produced variants (e.g., easily pronounced pronoun forms in Cantonese), and hearers may prefer easily perceived or salient variants (Labov, 1994; Shen, 1997). Combining speaker's and hearer's preferences with social constraints due to network structures can better understand the mutual influence of individual learning (e.g., who introduce the preference for variants) and social networks (e.g., structural factors restricting linguistic interactions among individuals) on diffusion. In this paper, we evaluate such influence based on a diffusion model and a number of typical network structures. In the rest of the paper, we describe the diffusion model and adopted networks (Sec. 2), show (Sec. 3) and discuss (Sec. 4) the results, and finally conclude (Sec. 5).

2. Pólya Urn Diffusion Model and Network Structures

The Pólya urn model was first designed to study contagion (Johnson & Kotz, 1977; Marshall & Olkin, 1993). In the original model, a Pólya urn is an urn containing red or green balls; at each timestep, a ball is randomly drawn from

the urn, and returned to it together with some balls having the same or different color. This process is then iterated. Our diffusion model, inspired by this prototypical model, contains N agents (individuals), each denoted by an urn. An agent is initiated with V tokens, each having a feature x_i (all feature values form F) and belonging to a particular type (v_1, v_2, … v_v). At each timestep, an interaction takes place between two or more agents, where a token v_i is drawn randomly from an agent (*speaker*), and p_i (the prestige of v_i, all prestige values form P) tokens of the same type are added to the same or other agent(s) (*hearer(s)*). Such drawing and returning processes repeat themselves, causing the distribution of variant types in all agents to change over time. *Mutation* may occur if an added token has a different type from the drawn one.

From a linguistic perspective, token drawing corresponds to production, and token adding corresponds to perception and knowledge update. Each type of linguistic variants possesses a *feature value* (similar to F). One type of variants may have a higher *prestige value* (similar to P) than the other, thus being preferred by speakers or hearers. During interactions, mutation corresponds to *transmission error*. Fig. 1 shows three examples of the Pólya urn dynamics in this model. This stripped-down model simulates production, perception, and update of variants during linguistic interactions, introduces variant preference from speakers or hearers, and traces the diffusion of variants in the population based on the distribution of different types of variants.

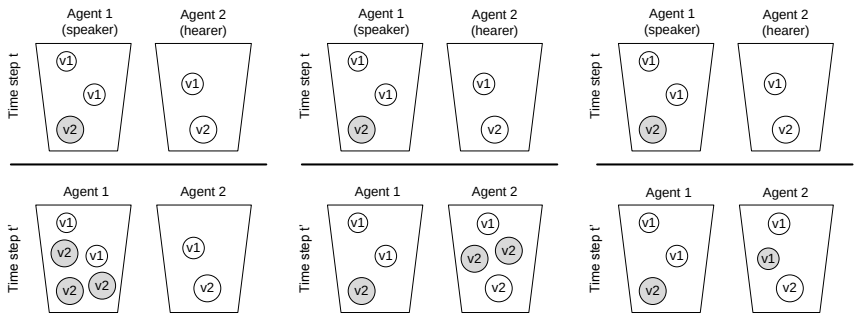

Figure 1. Pólya urn dynamics (N=2, V=2, F={1, 2}, P={1, 2}). Left: with speaker's preference: at time t, the speaker selects a token of v_2 to produce, and adds two tokens of v_2 to its own urn, which form the state at time t'; Middle: with hearer's preference, the hearer adds two v_2 to its own urn; Right: with hearer's preference and mutation, when the hearer adds two v_2, one is mutated to v_1.

The model uses social networks to define connections among agents, treating agents as nodes and connections as edges. It considers 6 types of representative networks: fully-connected network, star network, scale-free

network, small-world network, two-dimensional (2D) lattice, and ring. Table 1 lists the average degree, clustering coefficient and average shortest path length of these networks. It is shown that: from ring to 2D lattice, average degree increases; from 2D lattice to small-world network or scale-free network, average shortest path length drops due to 'short-cuts' in small-world network or 'hubs' (nodes having many connections with others) in scale-free network; from 2D lattice to scale-free network, and then, to star network, level of centrality increases; more and more nodes tend to connect to some popular node(s).

Table 1. Network characteristics: values are calculated based on 100 nodes. A scale-free network is formed by preferential attachment, with average degree around 4; a small-world network is formed by rewiring from a 2D lattice, with the reviewing rate as 0.1. Numbers within brackets are standard deviations of values in scale-free and small-world networks. *Average degree*: average number of edges per node; *clustering coefficient*: probability for *neighbors* (directly connected nodes) of a node to be neighbors themselves; and *average shortest path length*: average smallest number of edges, via which any two nodes in the network can connect to each other.

Network	Average degree	Clustering coefficient	Shortest path length
Fully-connected	99	1.0	1
Star	1.98	0.0	1.98
Scale-free	3.94 (4e-14)	0.14 (0.038)	3.01 (0.071)
Small-world	4	0.17 (0.031)	3.79 (0.086)
2D lattice	4	0.5	12.88
Ring	2	0.0	25.25

For the sake of simplicity, in the following simulations, we set N=100, V=2, F={1,2}, P={1,2}, 5000 (50 interactions per agent) pair-wise interactions (only agents that are directly connected can interact, each agent in a connection can be either speaker or hearer, and each connection in a network has an equal chance to be chosen for interaction), and exclude mutation. Among the two types of variants, Type 2 is preferred, either by speakers (speaker's preference) or by hearers (hearer's preference). In a simulation, we measure *proportion*(v_i,*t)* as the proportion of v_i in the population and *prop(t)* (=*max*{*proportion*(v_i,*t)*}) as the proportion of the majority type, at 50 sampling points evenly distributed among 5000 interactions. In each type of network, we conduct 100 runs with speaker's preference and 100 with hearer's preference.

3. Simulation Results and Statistical Analysis

We conduct a 2-way analysis of covariance (ANCOVA) (dependent variable: *prop(t)* over 100 simulations; fixed factors: speaker's/hearer's preference and networks; covariate: 50 sampling points). Using ANCOVA, instead of ANOVA at a particular sampling point, is to neutralize the influence of the number of

interactions. This analysis shows significant main effects of networks ($F(5, 61187) = 1111.425$, $p < .001$, $\eta_p^2 = .083$) and speaker/hearer's preference ($F(1, 61187) = 6905.606$, $p < .001$, $\eta_p^2 = .101$) on *prop(t)*. These effects are shown in Fig. 2, the left panel of which shows the marginal mean *prop(t)* with speaker's (solid line) and hearer's (dashed line) preference in different types of network (calculated as the mean value over all *prop(t)* at every sampling points) and the right panel shows the marginal mean *prop(t)* among networks (calculated as the mean value over all *prop(t)* in simulations with speaker's preference and with hearer's preference). Apart from these, the number of interactions has a significant effect ($F(1, 61187) = 108285.542$, $p < .001$, $\eta_p^2 = .639$).

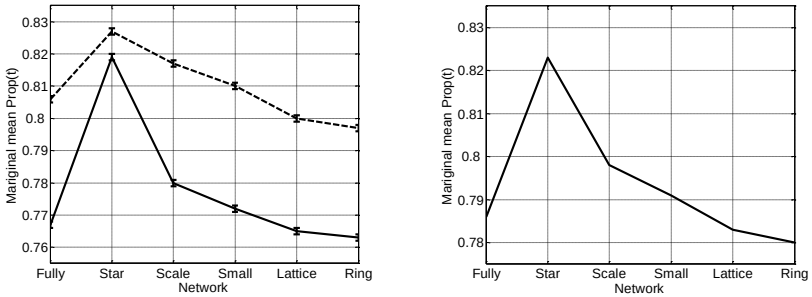

Figure 2. Simulation results. Left: marginal mean *prop(t)* with speaker's (solid line) and hearer's (dashed line) preferences in different networks. Right: marginal mean *prop(t)* in different networks.

In these simulations, due to its higher prestige value, Type 2 variants gradually become the majority type and diffuse in the population, indicated by high *prop(t)*. Meanwhile, speaker's and hearer's preferences show different diffusion efficiencies. This is evident in fully-connected network, which resembles the case of random interaction and excludes network influence. In this network, hearer's preference leads to a higher degree of diffusion than speaker's preference. Similar tendencies can be found in other networks. Furthermore, degrees of diffusion in different networks are different; star network has the highest *prop(t)* (the highest degree of diffusion), then, scale-free network, small-world network, lattice, and finally, ring.

4. Discussions

During one interaction, the preference introduced by a speaker or by a hearer has the same effect on the distribution of variants in these two contacting agents, as shown in Fig. 1. However, in a situation of multiple agents and iterated interactions, these two types of preference start to show different effects.

Speaker's preference is *self-centered*, disregarding other agents. For example, if an agent has a majority of Type 1 variants, when it interacts as the speaker with another agent whose majority variant type is Type 2, it still has a higher chance of choosing a token of Type 1 and increasing Type 1's proportion by adding tokens of the same type. In contrast, hearer's preference is *other-centered*, allowing the hearer's variant distribution to be adjusted by other agents. For example, if an agent has a majority of Type 1 variants, when it interacts as the hearer with another agent whose majority variant type is Type 2, it will have a higher chance of adding some Type 2 tokens, which will gradually adjust its variant distribution to become similar to others'. After iterated interactions, a uniform majority variant type in all agents can be achieved. Therefore, given the same number of interactions, hearer's preference for certain type of variants is more efficient for diffusion than speaker's preference, as shown by the significantly higher *prop(t)* with hearer's preference than that with speaker's preference in Fig. 2. Considering one-speaker-multiple-hearer interactions in real societies, the effect of hearer's preference will be further improved.

Due to the clear prestige advantage of Type 2 over Type 1, the value difference of *prop(t)* in these two preference conditions is not high, but the statistical analysis confirms that they are significantly different. This indicates that in real case diffusion studies we need to separate these two types of preference and pay attention to their possible different effects on diffusion, especially in a multi-agent human society.

The effect of networks on degree of diffusion can be ascribed to structural features of networks. The first candidate feature is *average degree*. Seen from Table 1, average degree is 2 in ring, 4 in 2D lattice. But the marginal mean values of *prop(t)* in these two networks are not significantly different (a post-hoc T-test of the marginal *prop(t)* does not show a significant difference, $t(198) = -0.759$, $p = 0.449$). However, in one-speaker-multiple-hearer interactions as in real societies, the effect of average degree becomes explicit (the same T-test shows a significant difference, $t(198) = -9.338$, $p < 0.001$). In addition, the similar results in ring and 2D lattice but different results in 2D lattice and scale-free or small-world networks indicate that there must be other structural features that are taking effect. And average degree alone fails to explain why star network, having the lowest average degree (1.98), has the highest *prop(t)*.

The second candidate is *level of centrality*. Star network has an extremely centralized structure; there is a hub connecting all the other nodes and participating in all interactions. With speaker's preference, the hub has many chances to update its variant distribution; with hearer's preference, any update

of variant distribution can be quickly spread via the hub to others. Besides star network, scale-free network, due to preferential attachment, also has several nodes that connect many others, but the level of centrality in scale-free network is less than that of star network. Accordingly, its *prop(t)* is significantly smaller ($t(198) = -5.194$, $p < 0.001$). Lacking centralized structures in other types of networks causes their *prop(t)* to be much smaller than that in star or scale-free network. Level of centrality also explains why small-world network and 2D lattice, having identical average degree, end up having significantly different *prop(t)*($t(198) = 3.538$, $p < 0.001$). Small-world network is created via rewiring from 2D lattice. Such rewiring causes some nodes to have more edges than others, and these nodes will play a similar role as hubs to increase *prop(t)*. However, this way of introducing hubs is less efficient than preferential attachment, so *prop(t)* in small-world network is significantly smaller than that in scale-free network ($t(198) = -3.878$, $p < 0.001$).

Finally, *average shortest path length* reveals a combined effect of average degree and level of centrality; if a network has a high average degree and/or a high level of centrality, any two nodes in it can be connected via a small number of edges. Since star network has much lower average shortest path length (1.98), it has much higher *prop(t)*, and then scale-free network (3.01), small-world network (3.79), and 2D lattice (12.88). Since ring has the highest average shortest path length (25.25), its *prop(t)* is the lowest.

5. Conclusions and Future Work

This study discusses the effects of speaker's and hearer's preferences for certain types of variants and those of network structures on diffusion of linguistic variants in a population of individuals. The distinction of speaker's and hearer's preferences and the discovery of their different effects on diffusion extend the previous work that focuses more on hearer's effect on language change or diffusion. Identification of the effect of level of centrality on diffusion enriches the discussion of socio-cultural factors relevant for language evolution. Finally, this study can trigger some promising future work, such as examining the diffusion with contradictory speaker's and hearer's preferences or analyzing the effect of preference introduced by crucial nodes (e.g., hubs) in networks, especially in one-speaker-multiple-hearer interactions.

Acknowledgements

Gong is supported by the Society of Scholars in the University of Hong Kong. Tamariz is supported by the UK AHRC Grant AH/F017677/1.

References

Abram, D. M., & Strogatz, S. H. (2003). Modeling the dynamics of language death. *Nature*, 424, 900.

Cheng, R. L. (1987). Borrowing and internal development in lexical change. *Journal of Chinese Linguistics*, 15, 105-131.

Croft, W. (2000). *Explaining language change: An evolutionary approach*. Harlow, UK: Longman.

Dall'Asta, L., Baronchelli, A., Barrat, A., & Loreto, V. (2006). Non-equilibrium dynamics of language games on complex networks. *Physics Review E*, 74, 036105.

Fisiak, J. (Ed.) (1995). *Linguistic change under contact conditions*. New York: M. de Gruyter.

Gong, T. (2011). Simulating the coevolution of compositionality and word order regularity. *Interaction Studies*, 12(1): 63–106.

Hruschka, D. J. et al. (2009). Building social cognitive models of language change. *TRENDS in Cognitive Science*, 13, 464-469.

Johnson, N. L., & Kotz, S. (1977). *Urn models and their applications*. New York: Wiley.

Ke, J-Y., Gong, T., & Wang, W. S-Y. (2008). Language change and social networks. *Communications in Computational Physics*, 3, 935-949.

Labov, W. (2001). *Principles of linguistic change: Social factors*. Oxford, UK: Blackwell.

Marshall, A. W., & Olkin, I. (1993). Bivariate life distributions from Pólya urn model for contagion. *Journal of Applied Probability*, 30, 497-508.

Minett, J. W., & Wang, W. S-Y. (2008). Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*, 118, 19-45.

Mufwene, S. S. (2008). What do creoles and pidgins tell us about the evolution of language? In S. S. Mufwene (Ed.), *Language evolution: Contact, competition and change* (pp. 272-297). London: Continuum.

Nakamura, M., Hashimoto, T., & Tojo, S. (2007). Language change among 'memoryless learners' simulated in language dynamics equations. In C. Lyon, C. L. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language* (pp. 237-252). London: Springer.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, 108, 95-117.

Shen, Z-W. (1997). *Exploring the dynamic aspect of sound change*. Berkeley: University of California Press.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, 2, 319-332.

Wolfe, P. M. (1972). *Linguistic change and the Great Vowel Shift in English*. Berkeley: University of California Press.