

On the importance of population structure in computational models of language change

Abundance of research during the last 15 years has proven the relevancy of computational modeling to study both the evolution of the language faculty and the evolution of actual languages (Cangelosi & Parisi, 2002). In the present contribution, we will focus only on latter type of evolution and highlight the following paradox: while the reasons for the success of the computational approach to language change lie in the natural way it factors in ‘population’ (which is crucial for language change), most earlier computational models have assumed populations to be homogeneous. They have neglected ‘population structure’ which has to do with the internal heterogeneity of populations on many levels. We aim at demonstrating that integrating complex population structures in a computational model is not only feasible, but indeed necessary.

Insights from sociolinguistics (Labov, 1963, 1966) and studies of language speciation (Mufwene, 2001) have demonstrated that linguistic variation in a community is both the source of language change and the synchronic manifestation of the competition among variants. It is therefore inadequate to explain language change without invoking the population structure in which the change takes place. Because computational modeling of language change proceeds either by deriving equations describing the consequences of the linguistics interactions within a population of speakers (e.g. Niyogi, 2006), or directly by simulating those interactions (e.g. Hurford, 2003), it is a useful tool for the study of language change. Yet, most of literature on the computational modeling of language evolution ignore population structure, as if members of the population were likely to interact randomly with each other on an equal basis (Kirby & Hurford, 2002, Komarova, 2006, Nowak *et al.*, 2002, Steels & Kaplan, 1998, among many others). Some attempts have been made to consider spatially organized populations (e.g. Niyogi, 2002) but models proposing social organizations of the linguistic community have remained very limited in number: Nettle (1999) is a notable exception, while Lee *et al.* (2000) have structured populations with networks with properties known to differ from actual social networks properties (Watts & Strogatz, 1998). To determine whether there is any form implicit legitimacy for the lack of interest in population structures in computational modeling of language change (which would contradict the fact that linguists have demonstrated the importance of population structure; Labov, 1966, Milroy & Milroy, 1985), we examine the consistency of the dynamics shown by a simple model tested against several population structures.

We consider a model derived from Mufwene (2002): two linguistic variants, V_1 and V_2 , are in competition in a linguistic community. Oversimplifying things for the purposes of our demonstration, we assume that speakers use only one of the variants and that learn language from N other speakers, or teachers, among which k use V_1 and $N-k$ use V_2 . The exact value of k for a given speaker depends on the frequencies f_{V_1} and f_{V_2} of the variants in her/his social environment. We further assume that in cases where no bias favors the use of one of the competing variants, a learner will select the variant s/he will use with a probability proportional to the number of teachers using it: $P(V_1)=k/N$ and $P(V_2)=(N-k)/N$. In cases where there is bias, we introduce a parameter, $b \in [0,1]$, which modulates the probability of learning the disfavored variant: $P(V_1)=b \cdot k/N$ or $P(V_2)=b \cdot (N-k)/N$. In those cases, the selection by the speakers of one or the other variants is thus influenced both by their frequencies and by the bias induced either by linguistic causes (such as unmarkedness), or by social causes (search for prestige, claims for group membership, and so on). We apply this simple model to various types of population structures, from fully connected social networks (*i.e.* unstructured populations), to more complex and realistic community-based populations, and we propose analytic solutions and computer simulations of the dynamics of variant frequencies in those populations. Our results show that those dynamics are directly influenced by population structures and therefore imply that any conclusion drawn from computational models of language evolution that ignore the relevance of population structure needs some rethinking.

References:

- Cangelosi, A. and Parisi, D. 2002. *Computer Simulation: A New Scientific Approach to the Study of Language Evolution*. In A. Cangelosi and D. Parisi, eds, *Simulating the Evolution of Language*, pages 3--28. London: Springer Verlag.
- Hurford, J. 2003. *Why Synonymy is Rare: Fitness is in the Speaker*. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, eds, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL), Lecture Notes in Artificial Intelligence*, Vol. 2801, pp.442-451, Springer Verlag, Berlin, Heidelberg.
- Kirby, S. and Hurford, J. 2002. *The Emergence of Linguistic Structure: An overview of the Iterated Learning Model*. In A. Cangelosi and D. Parisi, eds, *Simulating the Evolution of Language*, pages 121--148. London: Springer Verlag.
- Komarova, N. 2006. *Population dynamics of human language: a complex system*. In *Frontiers of engineering: reports on leading-edge engineering from the 2005 symposium*, pages 89--98. The National Academy Press.
- Labov, W. 1963. *The social motivation of a sound change*. *Word*, 19, 273-309.
- Labov, W. 1966. *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Lee, Y., Collier, T. C., Taylor, C. E., and Stabler, E. P. 2005. *The role of population structure in language evolution*. In *Proceedings of the 10th International Symposium on Artificial Life and Robotics*.
- Milroy, J. and Milroy, L. 1985. *Linguistic change, social network and speaker innovation*. *Journal of Linguistics* 21, 339-84
- Mufwene, S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S. 2002. *Competition and selection in language evolution*. *Selection* 3.45-56.
- Nettle, D. 1999. *Using Social Impact Theory to simulate language change*. *Lingua*, 108(2-3):95—117.
- Niyogi, P. 2002. *Theories of cultural evolution and their application to language evolution*. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.
- Niyogi, P. (2006) *The Computational Nature of Language Learning and Evolution*. Cambridge, MA: MIT Press.
- Nowak, M, Komarova, N., and Niyogi, P. 2002. *Computational and evolutionary aspects of language*. *Nature*, 417:611—617.
- Steels, L. and Kaplan, F. *Stochasticity as a Source of Innovation in Language Games*. In Adami, C. and Belew, R. and Kitano, H. and Taylor, C., eds, *Proceedings of Artificial Life VI*, pages 368-376, Cambridge, MA, June 1998. The MIT Press.
- Watts, D. and Strogatz, S. 1998. *Collective dynamics of 'small-world' networks*, *Nature*, 393:440-442