SPONTANEOUS DIALECT FORMATION IN A POPULATION OF LOCALLY-ALIGNING AGENTS

RICHARD A BLYTHE, ALISTAIR H JONES, JESSICA RENTON
School of Physics and Astronomy, University of Edinburgh
Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK
r.a.blythe@ed.ac.uk, alistair.jones@kcl.ac.uk, s1003707@sms.ed.ac.uk

A language can be viewed as a shared set of conventions that relate meanings (functions) to signals (forms). These conventions however emerge from episodes of language learning and use between individual members of a group. One important characteristic of these interactions is a 'principle of density' advocated by Bloomfield (1933) who states that "every speaker is constantly adapting his speech habits to those of his interlocutors" (quoted by Labov, 2001, p19). By now the question of how local alignment on different conventions can develop into a convention shared by an entire group is well understood, both from modeling (Castellano, Fortunato, & Loreto, 2009; Smith, 2014) and experiment (Centola & Baronchelli, 2015).

A robust prediction of many models is that a state of global consensus—a single set of conventions for the whole group—is almost always reached (Castellano et al., 2009; Blythe, 2015). This therefore precludes the emergence of multiple dialects within a mixing population, by which we mean the existence of multiple equivalent forms for the same function (whether at the phonological, lexical or some other level of linguistic structure) being used by agents who interact. These dialects are clearly a feature of human language, and arguably exist also in certain bird, dolphin and other animal communication systems, perhaps because they have evolutionary utility in allowing kinship or group membership to be discriminated (Fitch, 2000).

There are a few counterexamples to the prediction of certain consensus. If biases that favor a different specific convention within each subgroup pre-exist, a correlation between group membership and linguistic behavior can stably persist (Kandler & Steele, 2008). In this case, dialect formation is not spontaneous: the dialects simply mirror a set of pre-existing biases whose origins remain mysterious. One mechanism that *is* spontaneous, in the sense that individuals have no *a priori* bias towards one behavior over another, arises when individuals preferentially interact with those group members who exhibit similar cultural behaviour (Axelrod, 1997). Then, the network of social influence adapts to reflect differences in linguistic behavior. Intuition suggests that the converse process, whereby stable differences in linguistic behavior stem from (potentially static) variation in the degree of influence from different speakers, should also occur. This situation corresponds more closely to what is typically understood to be implied by Bloomfield's principle of density (Labov, 2001), but so far coexistence of multiple dialects has been seen only in the extreme case of highly heterogenous network structures (Dall'Asta, Baronchelli, Barrat, & Loreto, 2006). Blythe and Croft (2012) have proposed a mechanism by which this might occur under more general conditions: namely, a situation whereby agents *acquire* a bias in favor of behavior that correlates with that of influential agents they have previously encountered.

In this work, we show that this mechanism does indeed allow for spontaneous dialect formation under quite general conditions. We demonstrate this in a computational model with two contrasting types of influence structure in the network. When the network is divided into groups within which agents strongly influence one another but between which influence is weak, small initial differences in behavior between the groups can be amplified into distinct stable dialects, as long as the disparity in influence between and within the groups is large. This is consistent with what was already observed by Dall'Asta et al. (2006). However, we find that multiple dialects also emerge when agents are embedded in space and interact only with their neighbors, even if the degree of influence between interacting agents is uniform.

A linear stability analysis can be applied to determine the conditions under which an instability to the emergence of multiple dialects is present; this also shows that the state with multiple dialects is stable to small perturbations. This includes the fluctuations that are a consequence of linguistic interactions comprising a finite amount of data or agents having a finite memory. This result stands in contrast to game-theoretic models where dialects are neutrally stable (Zollman, 2005). In the case where the state of consensus is also stable, and cannot be escaped if reached, it is inevitably the case that a state with multiple dialects will eventually be destroyed. However, we will argue that the fact that speakers are discrete entities with variable behavior implies that multiple dialects may persist for many generations. A further feature of this model is that the bias derived from a historical average of interlocutors' behavior, weighted by influence, can be reinterpreted as a bias in favor of the local majority convention. This *regularization of variation* is typically found to be present in experiments involving linguistic and non-linguistic variation (e.g. Hudson Kam & Newport, 2005; Reali & Griffiths, 2009; Smith & Wonnacott, 2010), and is suggested to derive from such purely cognitive factors as memory limitations or a prior expectation that the world is predictable. This work suggests that Bloomfield's principle of density provides alternative origin of such an effect, in which regularization is a side-effect of variation in social influence between speakers.

References

Axelrod, R. (1997). The dissemination of culture a model with local convergence and global polarization. *Journal of Conflict Resolution*, *41*, 203–26.

Bloomfield, L. (1933). *Language.* New York: Henry Holt.

Blythe, R. A. (2015). Hierarchy of scales in language dynamics. *European Physical Journal B*, *88*, 295.

Blythe, R. A., & Croft, W. (2012). S-curves and the mechanism of propagation in language change. *Language*, *88*, 269–304.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*, 591–646.

Centola, D., & Baronchelli, A. (2015). The spontaneous emergence of conventions: An experimental study of cultural evolution. *PNAS*, *112*, 1989–94.

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

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–67.

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*, 151–195.

Kandler, A., & Steele, J. (2008). Ecological models of language competition. *Biological Theory*, *3*, 164–73.

Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Oxford: Wiley-Blackwell.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–28.

Smith, A. D. M. (2014). Models of language evolution and change. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*, 281–93.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–9.

Zollman, K. J. S. (2005). Talking to neighbors: The evolution of regional meaning. *Philosophy of Science*, *72*, 69–85.