An agent-based modelling approach to wave-like diversification of language families

Frederik Hartmann[1]

[1]University of Konstanz

The computational modelling of linguistic diversification processes and linguistic descent has been a recurring sight in cladistical research (e.g. Ringe, Warnow, and Taylor, 2002; Gray and Atkinson, 2003; Heggarty, Maguire, and McMahon, 2010; Chang et al., 2015). In the cases of NeighborNet models and Bayesian phylogenetic algorithms, linguistic diversification has been modelled as tree-like developments. In recent years, calls for wave-model implementations have become more prominent (cf. François, 2015) but so far computational algorithms for this task have not been introduced to the wider research community.

To resolve this issue, I propose the use of agent-based models (ABMs) to simulate linguistic diversification which are emerging tools in population dynamics and sociogeographic research (e.g. Weidmann and Salehyan, 2013; Bhavnani et al., 2014). These models create a set of acting and interacting agents on a geographical surface and simulate their behaviour and development over time according to certain parameters. In the case of linguistic diversification, these models can simulate linguistic communities that interact, migrate and innovate to observe their diversification behaviour over time.

In an exploratory study, I show how such an ABM algorithm can be devised and investigate its application, behaviour, and limitations based on a toy example. The toy example is a randomly generated diversification process on a hypothetical geographical surface which yields a certain 'observed' end stage. Then the ABM is started where during a single run, different linguistic communities emerge and develop their own patterns and linguistic features the specific evolution of which can be observed. The simulation is then run several thousand times under different parameters and after each run, its final stage is compared with the observed patterns in the toy data. Afterwards, those runs are extracted from the model that most closely fit the actual distributions in the final stage of the toy data. The results from this procedure show that such simulation models yield promising results as they can reconstruct wave-like linguistic spread across both the temporal and geographic dimensions. If the simulated diversification process predominantly shows a certain pathway in order to fit the data, we can take this as an indication that this pathway could have played a role in the actual development process of the linguistic communities in question. Further, they can take into account linguistic contact for which tree models yield less reliable results.

Ultimately, the findings of this study show that these models fit well with simulated data and are a promising method that can subsequently be tested on real-world examples.

Keywords: Linguistic diversification, wave model, phylogenetics, agent-based models

References

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing (2014). "Group segregation and urban violence". In: *American Journal of Political Science* 58.1, pp. 226–245.

Chang, Will, Chundra Cathcart, David Hall, and Andrew Garrett (2015). "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis". In: *Language* 91.1, pp. 194–244.

François, Alexandre (2015). “Trees, waves and linkages: Models of language diversification”. In: *The Routledge handbook of historical linguistics*. Ed. by Claire Bowern and Bethwyn Evans. New York: Routledge, pp. 161–189.

Gray, Russell D. and Quentin D. Atkinson (2003). “Language-tree divergence times support the Anatolian theory of Indo-European origin”. In: *Nature* 426.6965, pp. 435–439.

Heggarty, Paul, Warren Maguire, and April McMahon (2010). “Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1559, pp. 3829–3843.

Ringe, Donald A., Tandy Warnow, and Ann Taylor (2002). “Indo-European and Computational Cladistics”. In: *Transactions of the Philological Society* 100.1, pp. 59–129.

Weidmann, Nils B. and Idean Salehyan (2013). “Violence and ethnic segregation: A computational model applied to Baghdad”. In: *International Studies Quarterly* 57.1, pp. 52–64.