

## **SIMULATING CREOLE AND DIALECT FORMATION**

ZENA M. HIRA

*Department of Life Sciences, Imperial College London,  
South Kensington Campus, London, SW7 2AZ, UK*

MARK BARTLETT

*Department of Computer Science, University of York,  
Heslington, York, YO10 5DD, UK*

Human languages are in a constant state of change. New words are constantly being invented and old ones lost. Words change their meanings and pronunciations over time. In the extreme case, whole new languages are created, either through the merging of unrelated languages or the splitting of one language into many variants. These processes are known as creolisation and dialect formation respectively. The former is perhaps best known from those creoles which originated in the period of European colonization and slave trade and which still exist today in the Caribbean. In contrast, the latter process can be observed in a great many number of dialects and languages worldwide, such as the descent of the Romance Languages from Latin.

The use of computer modelling to study aspects of the evolution of language has become established over the past couple of decades. In particular, the naming games introduced by Steels (1995) have been very influential. Using this model, Steels showed how a shared vocabulary can emerge in a group of agents through a series of conversations involving two agents. The more times a word is used for an object, the stronger the association between the object and the word will become. New words can be added and words that do not have communicative success are removed.

Subsequently, Steels (1997) performed a limited evaluation of the element of distance in his game. In that paper, the agents are divided in clusters and eventually develop a stable vocabulary within the cluster and at the same time

they become familiar with the words used in the other clusters. In effect they become bilingual. Our work extends this model to study what happens when clusters merge or divide, simulating the conditions in which creoles or dialects form respectively.

In order to introduce multiple populations into the model of Steels, we place agents in a two dimensional environment in which agents are more likely to talk to agents closer to them. Additionally, we allow for this environment to be split in half and ban all communication between the halves, creating two subpopulations which are linguistically isolated from each other.

Simulations of language contact were performed by beginning in a split environment and then, after allowing each group to interact for a specific amount of time, the entire population was left to interact together. For small populations, such simulations did indeed lead to the formation of “creole” languages, with words in the final combined language being taken from both the languages spoken by the initial subpopulations. However, as the groups of agents become larger the agents have trouble in converging to a single language. No name is dominant over the other but in the conversations about an object the agents end up using one of the two names developed in the groups but they can never decide which one they prefer. This may be seen as more analogous to the creation of bilingualism in the agents.

By reversing the earlier process, first allowing the whole population to interact before splitting it into two subpopulations, the divergence of a common language can be studied. The results observed are that the similarity of the resulting languages in the groups is strongly related to the amount of time the agents have spent together in a single group at the beginning. The longer the time they have spent the higher probability they have to end up using the same language.

### **References**

- Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, 2(3), 319-332.
- Steels, L. (1997) Language learning and language contact. In Daelemans, W., Van den Bosch, A. and Weijters, A., (Eds), *Proceedings of the workshop on Empirical Approaches to Language Acquisition*, (pp. 11-24). Prague.