# Exploring Lexical Complexity and Language Coherence with an Iterated Learning Model

**Mike Dowman (mdowman@ling.ed.ac.uk)**
Language, Evolution and Computation Research Unit, Theoretical and Applied Linguistics,
University of Edinburgh, AFB, 40 George Square, Edinburgh, UK

## Introduction

A computer model of the cultural evolution of language was used to investigate the factors that determine the complexity of human languages, and to discover under what circumstances coherent languages will emerge. The model used the iterated learning methodology, in which multiple agents, each representing a person, learn from one another, thus simulating the evolution of language over several generations of speakers.

Barr (2004) investigated whether people must have common knowledge of conventions in order for coherent languages to emerge, or whether they can arise simply as a result of repeated local interactions. Some researchers have suggested that shared linguistic conventions can emerge only if agents reason about other agents' knowledge, and use words that they believe other agents will know. Barr created an iterated learning model in which coherent languages emerged simply through pair-wise interactions between agents. However, his agents only learned a mapping between four words and four meanings, and they received feedback concerning whether communication was successful.

Belpaeme (2002) also demonstrated the emergence of coherent languages, this time in a model of color term evolution. His agents also received feedback about whether communication was successful, and they adjusted their vocabularies accordingly. Belpaeme controlled the precision with which his agents needed to discriminate colors, thereby determining the number of words needed.

In another model of color term evolution, Dowman (2003) showed that coherent languages can emerge even when agents receive no feedback about whether communication is successful, and when there is no functional pressure to develop color terms with any particular communicative ability. In Dowman's model, agents simply used whatever word they thought most likely to denote a given color, and other agents observed these color and color term pairs. The number of words in the emergent languages was approximately proportional to the average number of times that each agent spoke during its lifetime, indicating that language complexity was determined by frequency of use.

## An Iterated Learning Model

In order to investigate why Dowman's (2003) model produced such results, a new expression induction model was made in which the agents simply remembered word forms, but did not associate meanings with them. When speaking, agents would randomly choose from amongst those words that they had heard at least twice, except that occasionally they would make up a new word. At random intervals, agents would be replaced by new agents who did not know any words. In the languages emerging after 100 generations, the average similarity between the vocabularies of the adult agents was always in excess of 97%. This was true even when there was a large number of agents, so that most agents would not have interacted with each other. Figure 1 shows that the number of emergent words was directly proportional to the number of times that agents spoke during their lifetimes. The frequency with which new words were introduced had very little effect on these results, except when it was raised to a very high level.

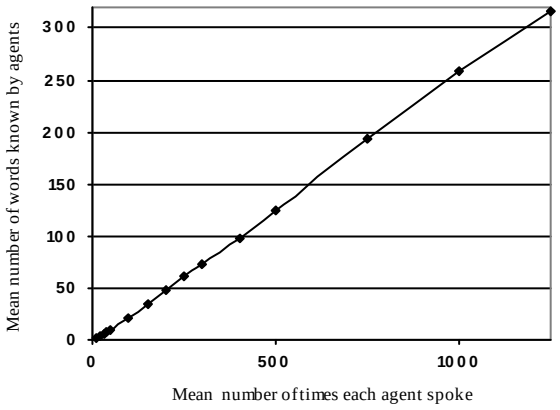


Figure 1: Relationship between frequency of use and number of words emerging.

These results show that coherent languages can emerge without feedback about communicative success, or consideration of other agents language knowledge. They suggest that language complexity may be controlled by how often language is used, rather than by any functional benefit arising from increased communicative ability.

## References

Barr, J. B. (2004) Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, *28*, 937-962.

Belpaeme, Tony (2002). *Factors influencing the origins of color categories.* PhD Thesis, Vrije Universiteit Brussel.

Dowman, M. (2003) Explaining Color Term Typology as the Product of Cultural Evolution using a Multi-agent Model. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society*.