Self-Organisation of a Learnable Language: What Happens When an Alien Language Replicates through Human Brains

Katrien Beuls, Vanessa Ferdinand, Jasmeen Kanwal

I. Introduction

Through the use of an iterated learning experiment, we investigate the emergence of self-organised structure in a random initial alien language, as it is replicated through successive 'generations' of language learners in three distinct iteration chains. Experiments in the iterated learning paradigm have frequently shown that languages evolve to become more compositional and thus more easily learnable and replicable over the course of an iteration chain (see, e.g., Kirby et al. 2008). Here, we closely analyse the dynamics of this replication, identifying which features actually correlate with greater replicability, as well as suggesting which mechanisms may be driving the evolution towards a more replicable language. These findings ultimately show us that certain very broad generalisations about language structure can be explained purely by appeal to the self-organisational dynamics of replication through the cognitive environment of individual language learners, even prior to taking social and cultural interaction and population dynamics into account.

II. Method

This experiment follows the iterated learning paradigm introduced in Kirby et al. (2008). In this paradigm, a subject is given a random initial 'alien language' to learn. The language consists of a set of randomly generated words paired with images which can be categorised according to a number of different features. The subject is trained on the initial set, and then asked in the testing phase to reproduce the names of the images. The forms they reproduce are then presented to the next subject to learn in the training phase, paired with the corresponding images, and the output from this next subject's testing phase are used as the input for the following subject, and so on.

In the present experiment, we randomly generated 15 words from the same syllables used in Kirby et al. (2008). While the latter study used these syllables to construct words of 2-4 syllables each, we instead used them to create words of length 4-6 syllables, with an equal number of words of each length. The extension of the word lengths in this experiment provided a means of testing whether there would be any significant reduction to average word length over successive iterations.

The 15 images paired with the words consisted of cartoon 'alien plants', which could be carved up into categories based on several different features, including shape, colour, texture, presence of thorns, and number of leaves. These feature classes ranged from binary to having 6 different possible values. In this way, our stimuli also differed importantly from that of Kirby et al. (2008), where the 27 images used made up an exhaustive set categorisable by 3 feature groups (shape, colour, and motion), each with 3 possible values. This alteration was made in order to test the hypothesis that only a subset of the possible features would end up being encoded in the final language, and to provide the opportunity of identifying which features these were. Some examples from the initialising set are reproduced in the figure above.

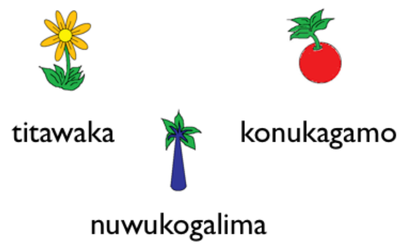

Fig. 1. Some example stimuli from the 'alien language' presented to the first subject of each chain.

Finally, another important variation in our method from that of previous iterated learning experiments is that we initialised our three different iteration chains (consisting of 12 subjects each) using the *same* starting stimuli set. This allowed for some novel ways of analysing the results of language transmission, such as tracking the divergence of the three chains from one another over time, as well as being able to compare which features of the initial set were most frequently preserved throughout all the chains.

III. Results and Discussion

Across all three iteration chains, we found that there were three main factors which correlated with higher replicability of a lexicon from one iteration to the next: a strong mapping between words/syllables and object categories (which resulted in fewer unique words, as only a subset of the possible object features were encoded, such as shape and colour); shorter average word length; and a smaller phoneme inventory. These were quantified using various different analyses, as shown in some example figures below (Figs. 2-4).

One notable result is that the three chains diverged significantly from one another over the course of the generations (see Fig. 5). This is illustrative of the range of possible trajectories that can result from the same initial conditions. It also lends further support to our hypothesis that shorter word length and

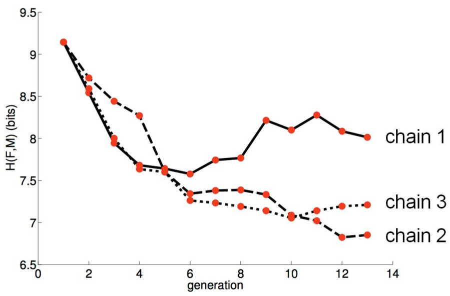

Fig. 2. The joint entropy of the bigram-meaning mapping distribution across the three chains. A higher entropy corresponds to a distribution where bigrams are randomly correlated with various object features, while a lower entropy implies a more strongly clustered distribution, where most bigrams occur consistently with a specific object feature.

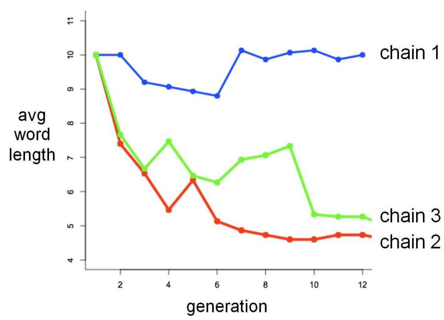

Fig. 3. Average word length in successive generations across all three chains.

stronger form-meaning mappings lead to higher replicability, as the chain with the longest average word length and the weakest form-meaning mappings (chain 1) had the lowest transmission fidelity, while chains 2 and 3 were transmitted fairly successfully towards the end.

But what mechanisms are causing the evolution in the alien language's structure, and why is this structure not always reached over the course of 12 generations? As shown in the computational models of Nowak and Krakauer (1999) and Plotkin and Nowak (2000), "errors during language acquisition increase the likelihood of reaching the optimum solution". In other words, it is the sensitivity to errors of the replication process through an individual's cognitive environment which

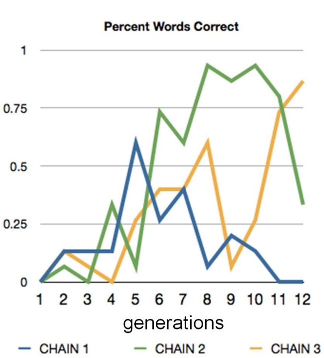

Fig. 4. Replicability of the language at each generation for all three chains.

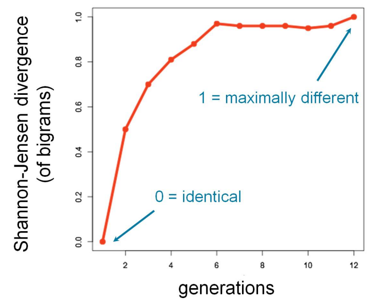

Fig. 5. The Shannon-Jensen divergence of the three chains over time.

allows for the possibility of both reaching more replicable 'attractor states' and of falling victim to perturbations away from such states. For example, in all three chains around 50% of bigrams were replicated in early generations, but they did not map consistently to specific meanings. However, given the finite meaning space, chance alignments of forms with meanings due to errors were highly probable, and new learners seemed to seize these statistical patterns and use them as evidence for genuine form-meaning mappings, consequently over-generalising the patterns in their output set. These same over-generalisation tendencies can also cause amplification of perturbations away from regular patterns, as seen in chain 1. This process bears a striking resemblance to genetic drift processes due to random sampling.

Acknowledgment

Thanks to Enrico Sandro Colizzi and Cameron Ray Smith for lively discussions and help with analysis, and to the Santa Fe Institute Complex Systems Summer School for providing the opportunity for this collaboration.

References

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105(31):10681–10686.

Nowak, M. A. and Krakauer, D. C. (1999). The evolution of language. *PNAS*, 96:80288033.

Plotkin, J. B. and Nowak, M. A. (2000). Language evolution and information theory. *Journal of Theoretical Biology*, 205:147–159.