Evolving an Artificial Creole

Gregory Furman, Geoff Nitschke
FRMGRE001@myuct.ac.za,gnitschke@cs.uct.ac.za
Department of Computer Science
University of Cape Town, South Africa

ACM Reference Format:
Gregory Furman, Geoff Nitschke. 2020. Evolving an Artificial Creole. In *Proceedings of The Genetic and Evolutionary Computation Conference 2020 (GECCO '20).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/n.n

EXTENDED ABSTRACT

There has been a significant amount of research on computational modeling of language evolution to understand the origins and evolution of communication [5–8]. However, there has relatively been relatively little computational modeling of environmental factors that enable the evolution of creole languages [2, 9], specifically, modeling lexical term transmission between intersecting language groups, within the context of artificial creole language evolution [4]. This study used an iterative agent-based *naming game* [8] simulation to investigate the impact of population size and lexical similarity [7] of interacting language groups on the evolution of an artificial creole lexicon. We applied the synthetic methodology [6], using agent-based artificial language evolution as an experimental platform to investigate two objectives. First, to investigate the impact of *population size* of interacting groups (with differing lexicons) on the evolution of a common (creole) lexicon. Second, to evaluate the concurrent impact of *lexical similarity* between interacting agent groups on the evolution of a creole lexicon.

Methods and Experiments

Experiments initialized a 40x40 bounded grid with a random distribution of 150 agents and 50 resources. With uniform randomness each agent was assigned a lexical *similarity threshold* value in the range: [0.0, 1.0], and an initial fitness value equal to 5. Each resource was assigned a value in the range: [1.0, 10.0], which corresponded to fitness received when agents consumed *that* resource. An agent's similarity threshold dictated what portion of a lexical term (character set) for a resource had to be from the given agent's own lexicon. Each agent was initialized with a lexicon of 26 *ASCII* characters (an agent's genotype representation). Experiments contained three agent types: *A*, *B* and *C*, where a given agent type was initialized with 26 randomly selected *ASCII* characters. All agents moved randomly about the grid for their *lifetime* (15000 simulation iterations), during which a variable number of naming games were played.

GECCO '20, July 8–12, 2020, Cancun, Mexico

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/n.n

A *naming game* was played whenever an agent moved to a grid position adjacent to a resource, and there was at least one other agent in the *talking radius*[1]. The agent then randomly selected one agent within the *talking radius* as a listener. The *two-agent naming game* then commenced with the first agent (*speaker*), sending a resource name of between 3 and 10 randomly characters (generated from its own lexicon), to the *listener*. The listener checked the word against its similarity threshold, to evaluate if the word contained a portion of characters (native to its lexicon) less than its similarity threshold. If so, the word was deemed *similar*, the naming-game terminated and this word was entered into the common (creole) lexicon. If the received word was deemed *not similar*, then the listener mutated (randomly changed a character to one from its own lexicon) or added a character (0.5 probability for either mutation, costing 1 fitness point). The speaker and listener roles then switched and the speaker (former listener) then spoke the mutated word back to the listener (former speaker). The listener then checked if the mutated word contained a portion of characters (from its native lexicon) less than its similarity threshold. This switching of speaker-listener roles and the naming game continued until the word was deemed similar by an agent or the the fitness of an agent dropped below a *minimum fitness threshold*.

When all agents adjacent to resources had played naming games, the next round of agent movements began. After 15000 agent movements, agents were assigned a fitness value equal to the resources they had consumed. One generation of artificial evolution equalled the population's lifetime (15000 iterations). Each generation, each agent was assigned a fitness equal to the resources consumed during its lifetime. *N-point crossover* [1] (where N was selected with uniform randomness between 1 and genotype length - 1) was then applied to the fittest 20% of agents, randomly paired, where portions of lexicons (genotypes) were swapped-over and child genotypes produced. The similarity threshold and lexical characters in child genotypes were then mutated (to another ASCII character) with 0.05 degree of probability. Similarity thresholds were mutated with *Gaussian mutation* ($\sigma = 0.5$) [1]. Sufficient child genotypes were produced to replace the least fit 20% of the population. Each experiment was 1000 generations and 50 runs and evaluated the impact of varying initial portions of different lexicons types (*A*, *B*, *C*), on the evolution of lexical similarity thresholds and thus agents adopting characters from other lexicon types. Experiments tested populations with agent type [*A*, *B*, *C*] lexical distributions: [A=50, B=50, C=50], [A=100, B=25, C=25], [A=25, B=100, C=25], [A=25, B=25, C=100]. Experiments computed average (over 50 runs) *similarity threshold* and *lexical proportion* (the portion of an agent's original lexicon, of type A, B or C, that entered the creole lexicon).

[1]Simulation and experiment parameters and source code is available online at: https://github.com/gregfurman/artificial-creole-2020

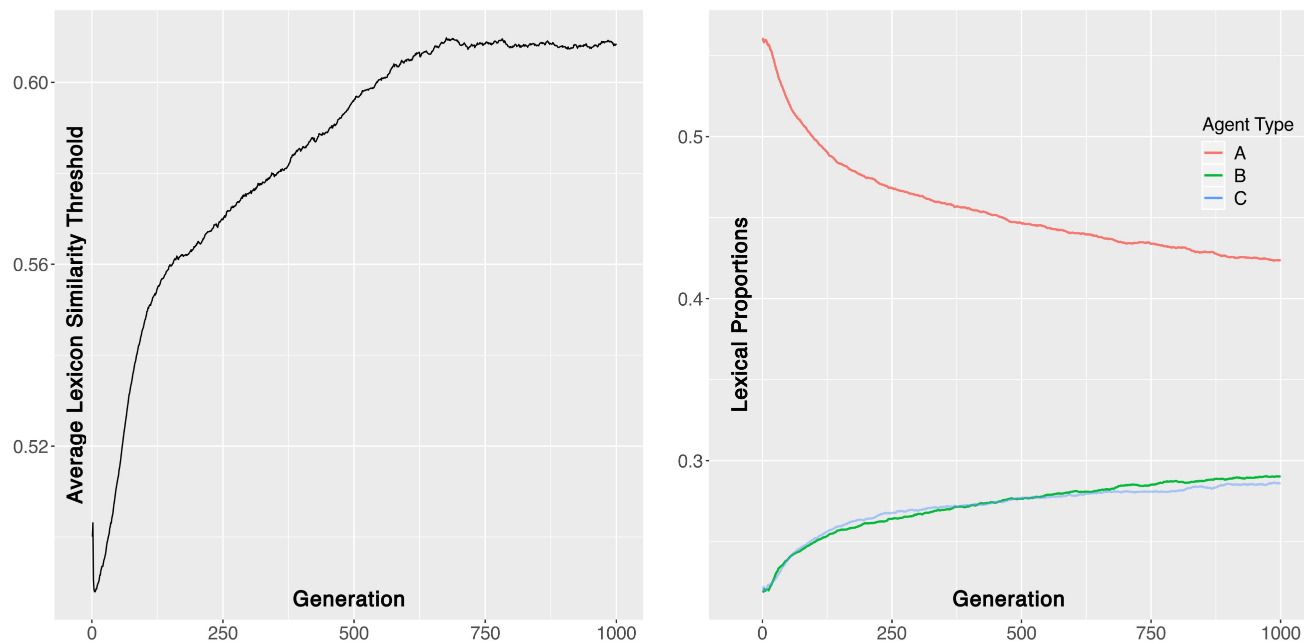

Figure 1: Left: Average population *similarity threshold*. Right: Average agent *lexical portions* over artificial evolution, given initial agent type (A, B, C) distribution: [A=100, B=25, C=25]. Averages were calculated over 50 evolutionary runs.

Results and Discussion

Results indicated for agent populations initially containing unequal portions of each agent (lexicon) type, the largest lexical portion (of the total population) decreases over evolutionary time and the minority lexical portions concurrently increase. This held for all experiments testing unequal agent type portions: [100, 25, 25], [25, 100, 25], [25, 25, 100]. As one example, figure 1 (right) illustrates this trend for the experiment testing the distribution: [100, 25, 25], for agent types A, B, C. Current research is testing whether this trend eventually results in convergence of statistically comparable portions of lexical types over multiple evolutionary runs. Given equal portions of lexical agent types, then the portion of character transfer between lexicons over evolutionary time was statistically insignificant (F-test, $p < 0.05$). Thus the initial portion of agents with a given initial lexicon type did not significantly impact the evolution of comparative lexicon type portions. This result is supported by related work in artificial creole evolution that similarly contends that common lexical structure is most impacted by the relative population sizes of intersecting language groups [4]. This trend of equalizing lexical portions is indicative of lexical merging over evolutionary time as different agents interact and agree on resource names to consume resources and gain fitness. Figure 1 (left) illustrates that in early-stage evolution the average (population) similarity threshold decreases to below 0.5 before eventually increasing to over 0.6 within 1000 generations. This is theorized to be a result of evolutionary selection for dissimilar lexicons maximizing lexical mixing and diversity in order to increase resource naming (expanding the creole lexicon) and thus boost fitness.

After a sufficient amount of lexical mixing has occurred, agents with progressively higher lexicon similarity thresholds (more similar lexicons) are selected for, meaning that the population's average similarity threshold increases as agent lexicons become more similar over evolutionary time. However, the evolutionary and environmental mechanisms responsible for this result remain the topic of ongoing research. To elucidate environmental factors influencing lexical transmission, mutation and merging of lexicons in natural creole evolution [3], current research is investigating the impact of factors including socio-economic influence of populations with initially distinct lexicons and varying lexical morphology similarity between interacting agent populations.

REFERENCES

[1] A. Eiben and J. Smith. *Introduction to Evolutionary Computing (2nd Edition)*. Springer, Berlin, Germany, 2015.
[2] F. Jansson, M. Parkvall, and P. Strimling. Modeling the evolution of creoles. *Language Dynamics and Change*, 5(1):1–51, 2015.
[3] S. Mufwene. The evolution of language: Hints from creoles and pidgins. In *Language Evolution and the Brain*, pages 1–33. City University of Hong Kong Press, Hong Kong.
[4] M. Nakamura and S. Tojo. The Emergence of Artificial Creole by the EM Algorithm. In *Proceedings of the International Conference on Discovery Science*, pages 374–381, Berlin, Germany, 2002. Springer.
[5] M. Nowak, N. Komarova, and P. Niyogi. Computational and evolutionary aspects of language. *Nature*, 417(1):611–617, 2002.
[6] L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1(1):1–34, 1997.
[7] L. Steels. Modeling the cultural evolution of language. *Physics of Life Reviews*, 8(1):339–356, 2011.
[8] L. Steels and A. McIntyre. Spatially distributed naming games. *Advances in Complex Systems*, 1(4):301–323, 1999.
[9] F. Tria, V. Servedio, S. Mufwene, and V. Loreto. Modeling the emergence of contact languages. *PLoS ONE*, 10(4):e0120771, 2015.