The Role of Speaker Prestige in Synthetic Language Evolution

Greg Furman\(^1\) and Geoff Nitschke\(^1\)

\(^1\)Computer Science Department, University of Cape Town, South Africa
FRMGRE001@myuct.ac.za, gnitschke@cs.uct.ac.za

Extended Abstract

Many researchers hypothesize that language adaptation, as with other evolutionary processes, entails both directed selection and random drift (Sapir, 1921; McMahon, 1994; Croft, 2000; Baxter et al., 2006; Van de Velde, 2014; Steels and Szathmáry, 2018). However, the specific contributions of these processes to language evolution remains an open question. It is well established that language evolution is not necessarily driven by selection, for example, speakers preferring specific word variants (Andersen, 1987; Blythe, 2012; Hamilton et al., 2016; Newberry et al., 2017).

Extending related work (Kandler et al., 2017), we use computational agent-based models to elucidate the impact of individual-level bias (speaker prestige) on population-level dynamics (average word similarity), where word diversity is measured by Levenshtein similarity (Levenshtein, 1966). Agents interacted in iterative language games (Kirby et al., 2014), to name and thus converse about resource types (A, B). Such object types represented conversation topics (Karjus et al., 2020c), where resource value indicated agent bias for conversing about (evolving words for) popular topics. For a null model comparison, we comparatively evaluated random drift versus directed word evolution on evolving word similarity, where using directed evolution, agent bias for adopting specific words (about resource types) increased with speaker agent social prestige (fitness).

While previous work has demonstrated selective advantages of various forms of speaker sociolinguistic prestige including competing word variants and borrowed words (Abrams and Strogatz, 2003; Labov, 2011; J. Hernández-Campoys and J. Conde-Silvestre, 2012; Kauhanen, 2017; Caludef et al., 2017; Monaghan and Roberts, 2019; Karjus et al., 2020a,c), there has been little research on the impact of speaker prestige on word diversity in language evolution.

Methods and Experiments

Experiments used random distributions of agent-resource combinations scattered in \(Q \times Q\) bounded grid worlds. Each combination was assigned five random environments with each environment randomly reset and re-run 20 times. Simulation parameters were: agent populations of 100-500 in increments of 100, resource amounts of 500, 1000, 2500, and \(Q \times Q = 50 \times 50, 75 \times 75,\) and \(100 \times 100\).

Resources were type A (50\%) and type B (50\%), with payout of 10.0 (type A, popular topics) or 1.0 (type B, obscure topics). Agents were initialised (iteration 0) with fitness of 10.0 and assigned random strings of 3-9 ASCII characters (words) for each resource type (A, B) in the environment. Agents moved about the grid randomly for 2000 iterations during which a variable number of evolutionary or random-Drift naming games were played. A naming game started when an agent moved atop a resource and at least one other agent was within the agent’s neighbourhood (adjacent cells).

Evolutionary naming games used Artificial Neural Network (ANN) agent controllers with a static layer of eight inputs and one output, where NEAT (Stanley and Miikkulainen, 2002) evolved hidden-layer connectivity and weights. Each agent’s ANN controller input all surrounding grid-cell information including: agents’ terms for resource types, agents’ fitness and potential resource payout, to determine a bid value to output. The highest bidding agent consumed the (talked about) resource (receiving payout), and the bid value was deducted from this winning agent’s fitness. All other agents adopted the winning agent’s word for the talked about resource type. Random-Drift naming games assigned a random agent’s word for a given resource to all others in the naming game with uniform probability.

Evolutionary (Sets 1-3) and Random-Drift (Set 4) experiments tested the impact of agent bias (prestige equated with fitness) on evolving word diversity for resource types (conversation topics A and B, of varying popularity).

Set 1: Resources are initialised in random locations. ANN bidding for word adopted by \(k\) agents (in talking game).

Set 2: Type A resources were popular (payout = 10.0) and type B obscure (payout = 1.0). Agents used ANN bidding for word to be adopted by \(k\) agents (in talking game).

Set 3: As per Set 2: Except after 1000 iterations, resource type B became popular and type A became obscure.

Set 4: As per Set 1, except randomly selected agent had its word for resource type adopted by all (talking game) agents.
Results and Discussion

Results used normalised Levenshtein similarity (LVS) (Levenshtein, 1966) to measure linguistic distance between agent words (LVS=0.0 means words are most dissimilar, LVS=1.0 words are most similar). Ordinary least squares regression (Flannery et al., 1986) analysis indicated statistically significant positive relationships between resource amounts and average LVS (t-test, \(t=56.623, p<0.01\)), controlling for other independent variables. A statistically significant negative relationship between average LVS and population size (\(t=-116.196, p<0.01\)) and environment area (\(t=-5.235, p<0.01\)), was also observed. As was a statistically significant increase in average LVS from simulation iteration 0 to 2000 by an average of \(0.02667 (t=138.464, p<0.01)\).

Two-way ANOVA (Flannery et al., 1986) computed differences in mean LVS (iteration 2000) between experiment sets, population sizes, resource numbers, and environment areas (figure 1A-D), showing statistically significant differences (F-test, \(F=7.45, p<0.01\)) in at least one of the average LVS when comparing experiment sets. In post-hoc analysis, 2-Sample Kolmogorov Smirnov (Massey, 1951) tested if mean LVS per iteration per experiment were generated from the same distribution, indicating all experiment sets were generated from different distributions (\(p<0.01\)). Tukey's HSD test (Abdi and Williams, 2010) indicated statistically significant differences (\(p<0.01\)) in mean LVS of random-drift experiments (Set 4) and evolutionary experiments (Set 1-3), but with no significant difference (\(p>0.10\)) between the mean LVS of evolutionary sets (Set 1-3).

Results indicated that there was no significant difference in the average similarity (LVS) of words propagated in the population via evolutionary naming games (Sets 1-3, figure 1A). As in related work (Karjus et al., 2020b), individual-level bias (bidding in this study) resulted in increased word similarity at the population-level. This indicates the critical role of directed (evolutionary) word selection on population dynamics. Supporting this, figure 1(E, F) presents example clusters of similar words at the start (Figure 1F) versus end of evolution (Figure 1E), where statistically significant LVS differences between evolutionary (sets 1-3) and random-drift (set 4), experiments was observed (figure 1D). However, for all experiments, results indicated varying environment parameters (population size, environment area, resource amount) significantly impacted average LVS of words in the population. Average word similarity decreased with population and environment size (figure 1A), but increased with resource amounts (figure 1B), that is, increasing with the number of possible conversations.

Thus, larger population and environment sizes yield high diversity in the population’s words, where word diversity change is not driven by directed selection (Sindi and Rick, 2016; Newberry et al., 2017) (speaker prestige) or random drift (Reali and Griffiths, 2010; Blythe, 2012), but by the number of potential naming games (conversations). To further ascertain environmental impact on language evolution (Greenhill, 2016), ongoing research is investigating how individual-level cultural and social bias changes topic popularity (Karjus et al., 2020c) and social networks (Ke et al., 2008; Fagyal et al., 2010; Kauhanen, 2017), and impact on population-level dynamics such as corpus diversity.
References


