Aging and Language Change

Ellis Cain

Introduction

Aging

Language change

Model

Overview

Description

Dynamics

Results

Discussion

a) For my final project, I will be modifying the model from: Beeksma, M., de Vos, H., Claassen, T., Dijkstra, T., & van Kemenade, A. (2017). A Probabilistic Agent-Based Simulation for Community Level Language Change in Different Scenarios. *Computational Linguistics in the Netherlands Journal*, 7, 17-38.

Their original goal was to simulate historical language change using an agent-based model. They used word order change as a case study, where the model included two potential word orders and two groups, that initially started out with a certain proportion of using the different word orders. They manipulated the learning rate, likelihood of interaction, location-bound dialects, and agent variation within the population. For the outcome, they measured the proportion of the two word orders over time.

Their model consists of two populations of agents that all have individual language models, which influence their interactions. Each agent has a language model that determines their utterances; when they hear an utterance, it is added to their LM, which they draw from to produce their own utterances. Every step, two agents are selected at random to "interact" (speak and listen), which leads to them updating their language models.

Their main assumptions are as follows (paraphrased from their paper):

1. Language learning is based on imitating others.
2. Variability of an individual's language model is limited.
3. Language can be influenced by external factors (group membership, life expectancy, "norms").

I aim to explore how potential changes in learning rates across the lifespan may impact overall population-level patterns of language change. Based on the original paper, I have modified the assumptions of the model:

1. Language learning is based on imitating others, **though this may change over the lifespan and may depend on group membership**.
2. Variability of an individual's language model is limited.
3. Language can be influenced by external factors.

Currently, I have their original model running, though the outputted graphs differ slightly. The 1st major step is to **modify the model such that agents "age", which changes their learning/sensitivity parameters.** For example, agents may initially be more willing to learn / sensitive to change, but as they "age", this rate will decrease. The 2nd major step is to **modify the cross-interaction (between group interaction) such that learning will be dependent on group preference**. For example, young agents may be more likely to listen and learn from other young agents. The "stretch goal" is to **modify the model to represent meaning change (instead of word order).** Meaning change will be modeled similar to Kirby 2015, where language has forms {a, b, ...} and meanings {01, 02, 12, ...}. Agents will produce utterances that consist of a form and the intended meaning (i.e. "a01").

b) Similar to the model description above, the system decomposes into language and agents. The language consists of exemplars, which are all the possible utterances; in the original paper there were four possible combinations (two adverb types, two verb types), while the modified version will (potentially) consist of surface forms and meanings (see above).

Each agent has an language model that contains various exemplars, and has an initial preference towards using one type of construction. After my modifications, each agent will also have an age and group preference. As the agents age, their learning rates will decrease based on a specified factor. With group preference, agents may be less likely to listen and learn from out-group members.

Then, at each step, two agents are chosen at random to converse. One agent generates an utterance based on their internal language model, and the other agent hears and integrates that utterance into their language model (dependent on age and group preference).

c) The main outcome measure is the proportion of types of utterances over time. I will do batch runs that vary the influence of age and group preference on learning rates. For the influence of age, "uniform" will be the default, where the learning rate is kept constant regardless of age. Potential variations could be a linear decrease in learning rate as age increases, or maybe a sigmoid function. For the influence of group preference, I plan to initially keep it as a constant value. I think it will be implemented as out-group learning rate, such that the default will be when learning-rate == out-group learning rate. Then, I can vary instances where the equality varies, such that the out-group learning rate is higher or lower than the overall learning-rate. There will be another option, where this group preference (out-group learning rate) changes with age as well.

d) I think with the first two goals (age, group preference), it seems like an exciting project looking at the influence of aging on language change. I'm not quite sure about implementing the meaning version of the model (which is why I put it as a stretch goal). For the basic word order version, the outcome measure is similar to the SIR-type models, where the main interest is how a certain type of idea spreads through a system over time. Language change, then, is just the overall prevalence of each type of word order. For the meaning version, though, I would have to implement some "mutation" factor, where the least used surface form mutates its meaning (i.e., "a01" is used least, so it would mutate to "a02"). Then, outcome measure would need to be changed to track both surface form and meaning (something like most frequent meaning per surface form, amount of meaning change per surface form).