Teaching an old dog new tricks? Learning rates, aging, and language change

Ellis Cain

Abstract

Language change can be understood as representing the dynamic relationship between group-level and individual usage patterns, where individual usage influences future learning and production. While previous research has shown that language learning capabilities are present well into adulthood, along with age-related differences in language processing and representation, the impact of learning rates on language change is not well understood. The current research uses an agent-based model of (syntactic) language change to explore the interaction between learning rates, aging, and between-cohort communication with regards to the spread of language variants through a population. Namely, the results demonstrate that the time to equilibrium in the system is modulated by these factors, more than the final outcome of the system.

Introduction

Language changes over time

Which sounds more correct, "the snazzy snake *sneaked* through the snow-covered snacks" or "the snazzy snake *snuck* through the snow-covered snacks"? Based on an analysis of lexical usage in a diachronic text corpora, "sneaked" is being replaced by "snuck" in common usage (Michel et al., 2011). How do these dynamics of language change play out in a population?

Language can be viewed as a complex adaptive system, where it is a system of speakers in a speech community, whose behavior is based on past interactions, which then influence future behaviors ("Five Graces Group" et al., 2009). The dynamics of this complex adaptive system change over historical time, through rising and declining trends for words across the collective usage statistics of the collective (Bybee, 2015; Michel et al., 2011). Figure 1 below shows an example of this feedback loop between collective usage patterns and individual language usage. If language change is an emergent property of this feedback loop, how might learning impact language change?

Interactions between aging and language

Previous research has shown that the language learning ability is present well into adulthood (Brysbaert et al., 2014; Castro et al., 2021; Ryskin et al., 2017), while other studies have shown that cognitive capabilities may decline across the lifespan (Hartshorne & Germine, 2015; Salthouse, 2009). Still, the developmental patterns of associative networks across the lifespan may reflect changes in both cognitive control along with accumulation of lexical information (Dubossarsky et al., 2017). Therefore, given that there may be age-related differences in (language) learning abilities, how would changing learning rates, aging, and group membership impact the types of patterns observed in language change dynamics?

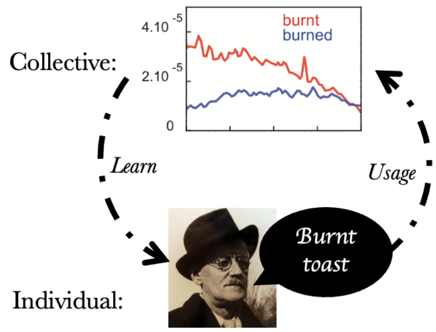

Figure 1: Example of the feedback loop in the complex adaptive system.

Model

Overview

The original language change model from Troutman et al. (2008) models the usage and spread of a grammatical variant throughout a population, such as a past tense ending ("-t" or "ed", as in "burnt" or "burned"). I modified the model such that each language user in the community has a learning rate that can be impacted by their age, which allows for an exploration of how changes in learning rates across the lifespan impact the overall collective-level patterns of language change.

The model's assumptions are as follows, modified from Beeksma et al. (2017) and Troutman et al. (2008).

1. Language learning is based on imitating others.
2. Learning rates may change over the lifespan.
3. There are variations between individuals.
4. Language can be influenced by external factors, such as group membership.
5. Language change has multiple stable equilibrium (*no set outcome*).

Here, equilibrium refers to when the community ceases to experience change in individuals' preference for one grammar or the other. For example, one equilibrium could be all language users only using Grammar 1, while another equilibrium could consist of half of the users using Grammar 1 and the other half using Grammar 0, which would correspond to a community with multiple dialect sub-groups.

With this model, the main outcome measures will be:

1. Final state of the system: *mean state of nodes in the population.*
2. Time to equilibrium: *the first time step that the group reaches that mean state.*

Code for the model and analyses can be found at: https://github.com/ellissc/aging_abm

Model description

The model initializes a preferential attachment network, where each node in the network represents a language user. There are two grammar variants, which are coded as 0 or 1, and are distributed throughout the

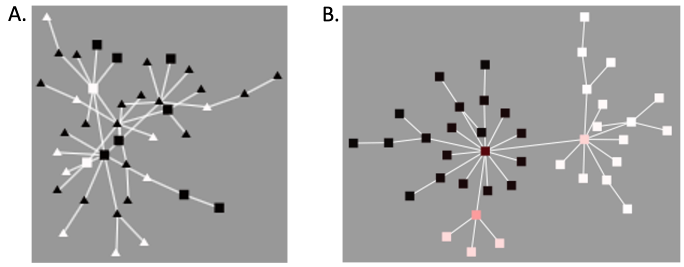

Figure 2: A) Example network generated in the model. B) Example of dialect sub-groups at the end of the simulation. Color represent a node's internal state, or bias towards one grammar or the other.

population based on a specified percentage. Figure 2A shows an example of the preferential attachment network.

Nodes have the following properties; state, spoken state, age, cohort, and gamma. *State* represents the node's current grammar state, ranging from [0,1], which influences their likelihood to use one of the two grammar variants. It can be understood as the bias towards a specific grammar (e.g., node with state of 0.25 would be biased towards Grammar 0 when speaking). *Spoken state* represents the output of each node's speech, as (grammar) 0 or 1.

The node's *age* should be self-explanatory, and it increases by 1 each tick. During initialization, the age of the nodes can be distributed according to a uniform or normal distribution. Likewise, there are two possible *cohort*s, which are initialized at the start. The cohort ages can be specified, such that Cohort 1 starts at *age = 0* and Cohort 2 starts at *age = 50*. Moreover, the grammars can be distributed at different percentages for each cohort, such that Cohort 1 starts with 25% using Grammar 1 and Cohort 2 starts with 50% using Grammar 1. Lastly, *willingness to listen to out-group* can be specified, such that for each "communication", a given node has a chance to ignore out-group members.

Gamma is the learning rate of a given node. Higher gamma values (0.08) mean that the node will adjust their state more when listening, as opposed to a lower gamma (0.01). It can be initialized according to a uniform or normal distribution, or it can be set to decrease with age.

Model dynamics

There are three main steps in the dynamics of the model: speaking, listening, and aging.

Regarding *speaking*, each node asynchronously generates an 'utterance', which consists of a discrete output (either 0 or 1, think of 'burnt' or 'burned'). This is based on a logistic curve that relates a node's internal state (weight of using Grammar 1) to the probability of using Grammar 1.

Regarding *listening*, all of the connected neighboring nodes will listen and use a linear reward-penalty algorithm to update their state (Bush & Mosteller, 1951, 1955; Yang et al., 2002). For this, they select a grammar that will be used to interpret the utterance (0 or 1). If the listener's selected grammar matches the heard utterance, they successfully interpret the utterance and update their internal state towards the heard state, otherwise, it will be updated away from the heard state. Here, the gamma parameter modifies the step size that the listener node takes when they update their internal state. As previously mentioned, if *willingness to listen to out-group* is specified, there may be a chance that a given node does not update their internal state if the speaker is a member of the other cohort.

Regarding *aging*, every node ages with each tick. If gamma is set to decrease with age, it decreases based on a "learning curve" function. This "learning curve" is defined using a power law relationship between *age* and *gamma* (learning rate), and is as follows:

$y = -0.05(x^a) + 0.005$

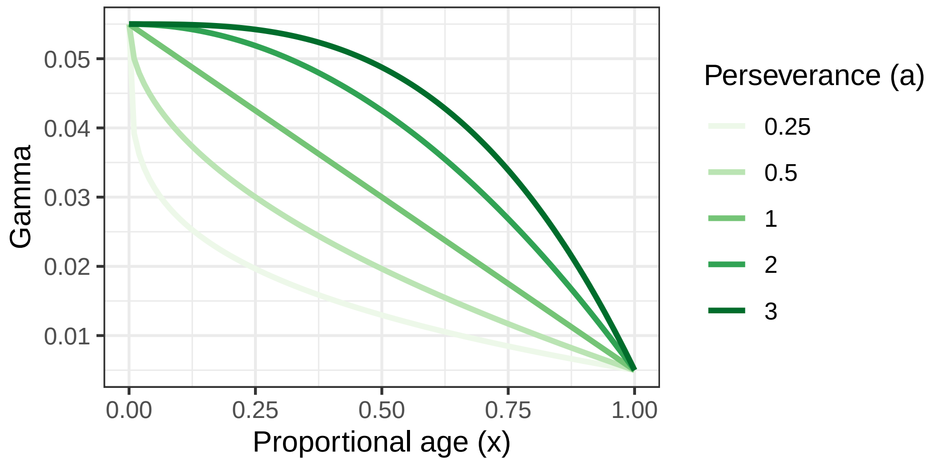

Figure 3: Graph of learning curve, which describes how learning rate may change with age. Colors represent different perseverance values.

The variable x is the normalized age of a given node, such that it is within [0,1]. The variable a represents the "perseverance" of learning, such that higher perseverance values ($a = 3$) indicate slow initial change that speeds up closer to x = 1, and lower perseverance values ($a = 0.25$) indicate a quick initial change that slows down later on. Graphs for the (decreasing) learning curves can be seen in Fig. 3.

Results

Each model configuration was run 50 times, with 1000 steps total per run. The networks were generated with 40 nodes, based on the original Troutman et al. paper. Using this agent-based model, I looked at the impact of learning rate and learning perseverance, aging and cohort age difference, and cohort anti-preference. Again, 'equilibrium' will refer to the stable state that the system reaches.

Impact of learning rate

Figure 4 shows the impact of learning rate on the final state of the networks. When the initial percentage of Grammar 1 is low ($<$ 35% of nodes), it rarely spreads through the population, as almost all of the runs end with grammar 0 as the majority. On the flip side, when Grammar 1 is present in the majority of agents ($>$ 60% of nodes), it will spread through the network the majority of the time. However, when it is present in half of the nodes, dialect subgroups will emerge (Fig. 2B), where different areas of the network will settle on different grammars.

For the variable that was manipulated between runs, gamma, there is no systematic impact on the final state that the system reaches. Regardless of how quickly nodes learn, the final result is mainly determined by the initial percentage of Grammar 1. However, when we look at the individual runs, we see a marked difference as gamma is manipulated. Figure 5 plots the mean state of the system over time for individual runs, based on various configurations of gamma and initial percent of Grammar 1. Namely, we see that while the equilibrium reached by the system does not differ between different gamma values, the speed at which it reaches that equilibrium decreases as gamma increases.

Figure 6 shows the proportional time to equilibrium for different values of gamma. The initial percentage of Grammar 1 does not greatly impact the time to equilibrium. When the learning rate is the slowest (gamma = 0.01), the system will reach equilibrium at a little less than 400 steps (0.4), while the fastest learning rate will cause the system to reach equilibrium in about 100 steps. As gamma increases, the change in time to equilibrium also decreases as well, seemingly approaching a limit of information transfer through the system.

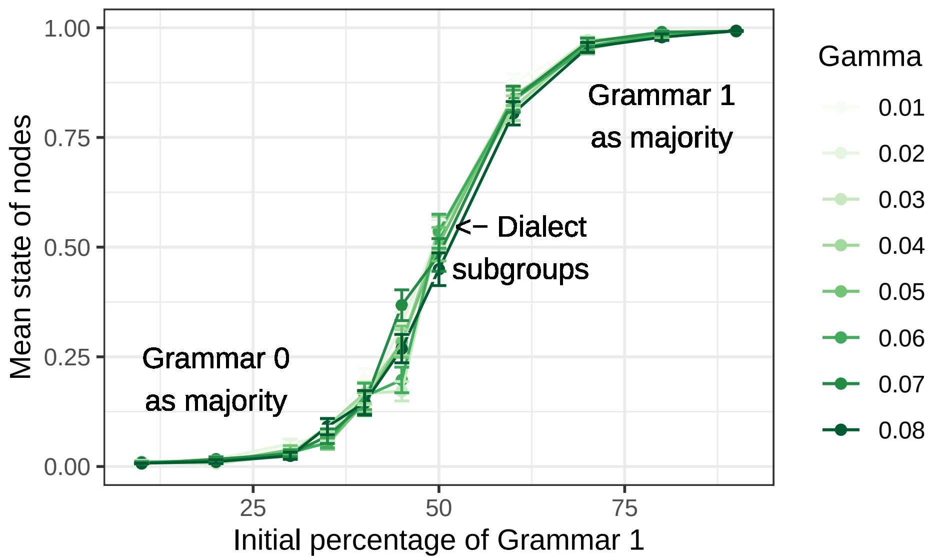

Figure 4: Final state of the langauge community, as measured by the mean state of nodes. Color of lines represents the gamma, or learning rate, of the nodes. Gamma is constant throughout the lifespan. Bars represent standard error.

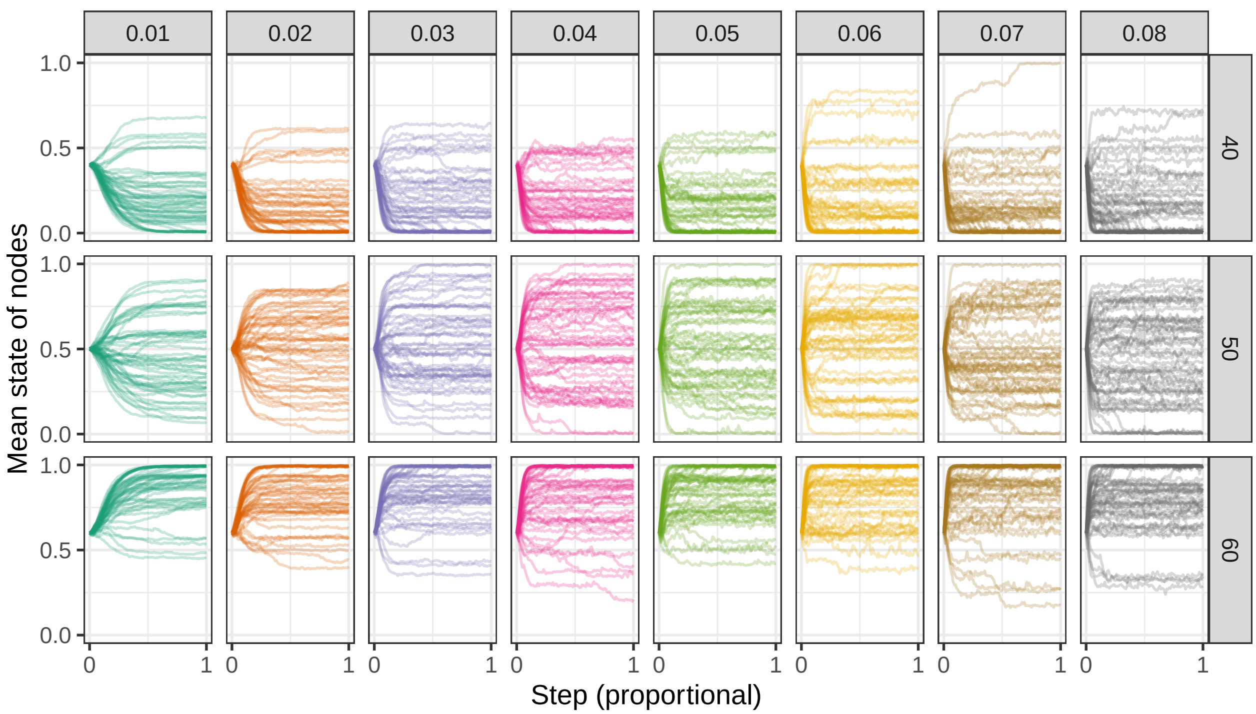

Figure 5: Each line represents the average state of the system for individual model runs. Here, the equilibria can be seen when the mean state flattens out. As gamma increases, the time to the equilibrium decreases. Rows faceted by initial percentage of Grammar 1, columns by the gamma value.

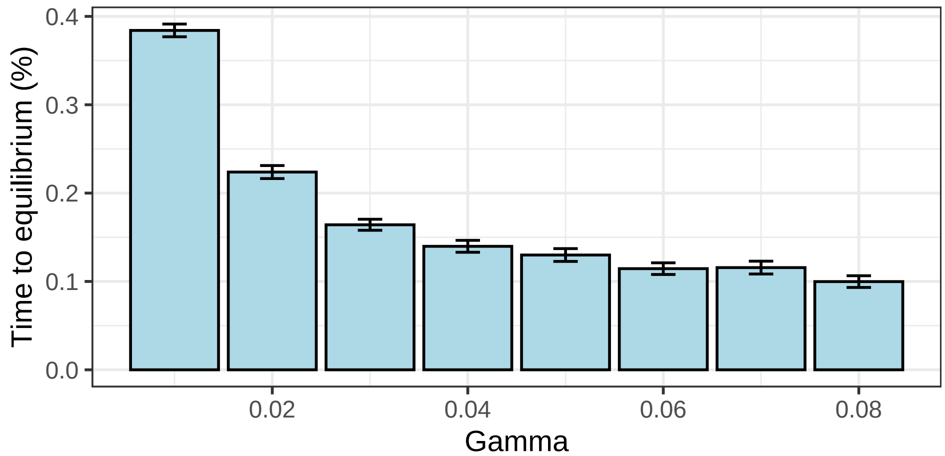

Figure 6: Time to equilibrium across the different values of gamma. Bars represent standard error.

In summary, gamma (learning rate) impacts the time to equilibrium of the system, and this effect size decreases as gamma increases.

Learning perseverance

For these next simulations, gamma will be determined by age, and the perseverance of learning (rate of decrease) will be manipulated. Refer to Fig. 3 for the relationship between age and gamma, where decreasing perseverance increases the speed at which learning rate decreases with age. Figure 7 shows the impact of learning perseverance on the mean state of the nodes at the end of the simulation. Similar to before, there does not seem to be a systematic relationship between perseverance and the final state.

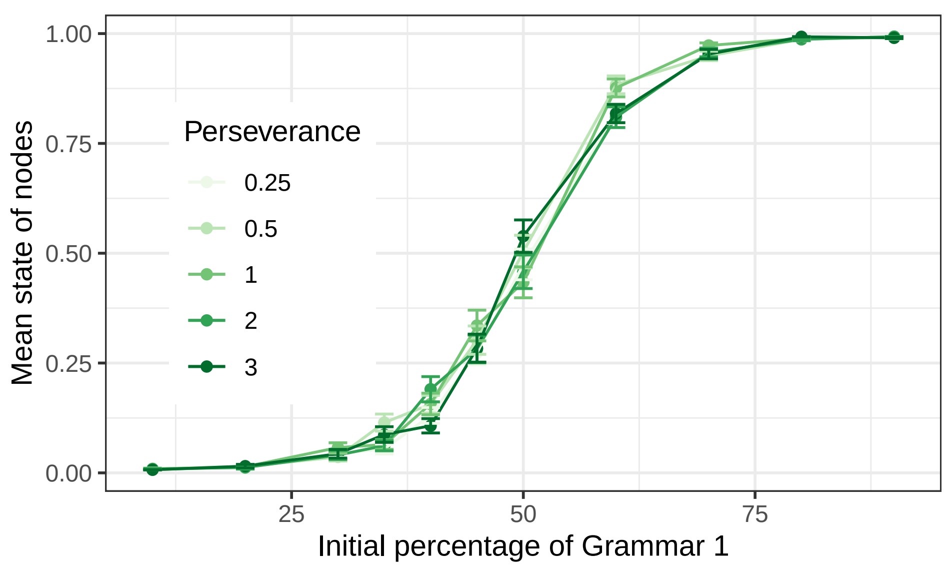

Figure 7: Impact of learning perseverance on the final outcome of the system.

However, when we plot the relationship between perseverance and time to equilibrium (Fig. 8), we see that most of the learning rate curves ($0.5 \leq$ perseverance ≤ 3), the time to equilibrium is very similar. Only when the learning rate quickly decreases (perseverance $= 0.25$) does the time to equilibrium increase. The range of time to equilibrium for these perseverance variations are similar in range to when gamma is varied.

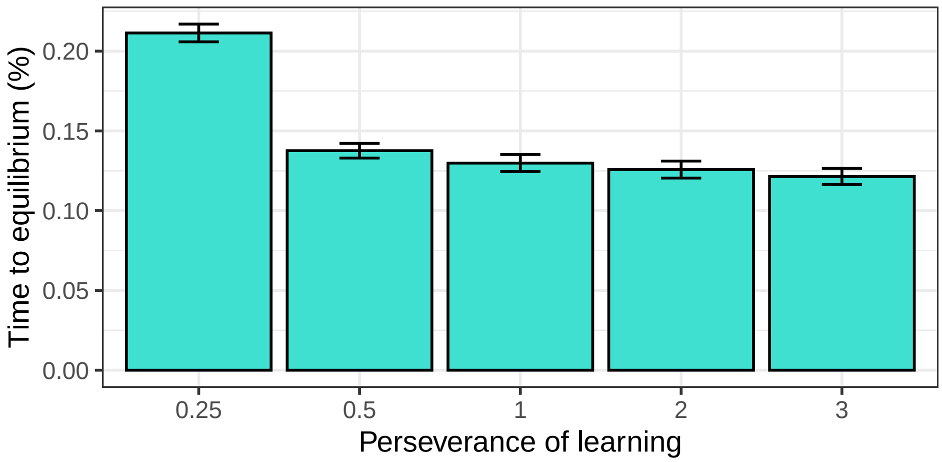

Figure 8: Time to equilibrium across the different values of perseverance, with standard error bars.

In summary, there is little difference of perseverance's impact on time to equilibrium when perseverance is greater than 0.5.

Aging and cohort equilibria

For this next set of comparisons, gamma will be set to decrease with age, with perseverance set at 1. In the simulations, there are two age cohorts, which the age gap between is varied. Similar to before, the initial starting percentage of Grammar 1 will be varied as well.

Figure 9 shows the time to equilibrium for different combinations of the initial percentage of Grammar 1 and population composition (percent of nodes in cohort 2). Cohort 1 started at age 0, so the different facets represent the age gap between the two cohorts. First, as the age difference between cohorts increases, the simulations where Grammar 1 was used initially by half the population increasingly takes longer on average to reach equilibrium. This effect generally seems to be more pronounced for the simulations where the majority of nodes were cohort 2. Then, when the initial percentage is at an extreme (10% or 90%), the time to equilibrium increases as well and has a larger impact on simulations with majority cohort 2.

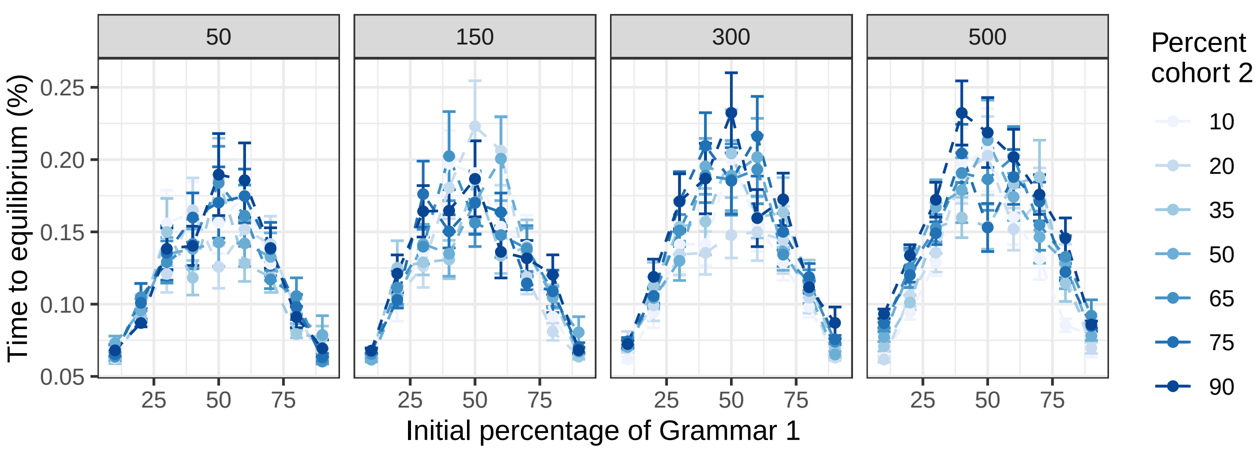

Figure 9: Time to equilibrium as a function of initial percentage of Grammar 1, across different population compositions. Faceted by cohort 2 age, which is the age gap between the two cohorts.

Figure 10 shows the separate equilibrium for each cohort, based on the initial percentage of cohort 2. The time to equilibrium seems to be generally consistent across the different variations, regardless of the age gap. Cohort 2, however, takes longer to reach equilibrium as they become more numerous in the population. Then, as the age gap increases, this impact on time to equilibrium increases as well.

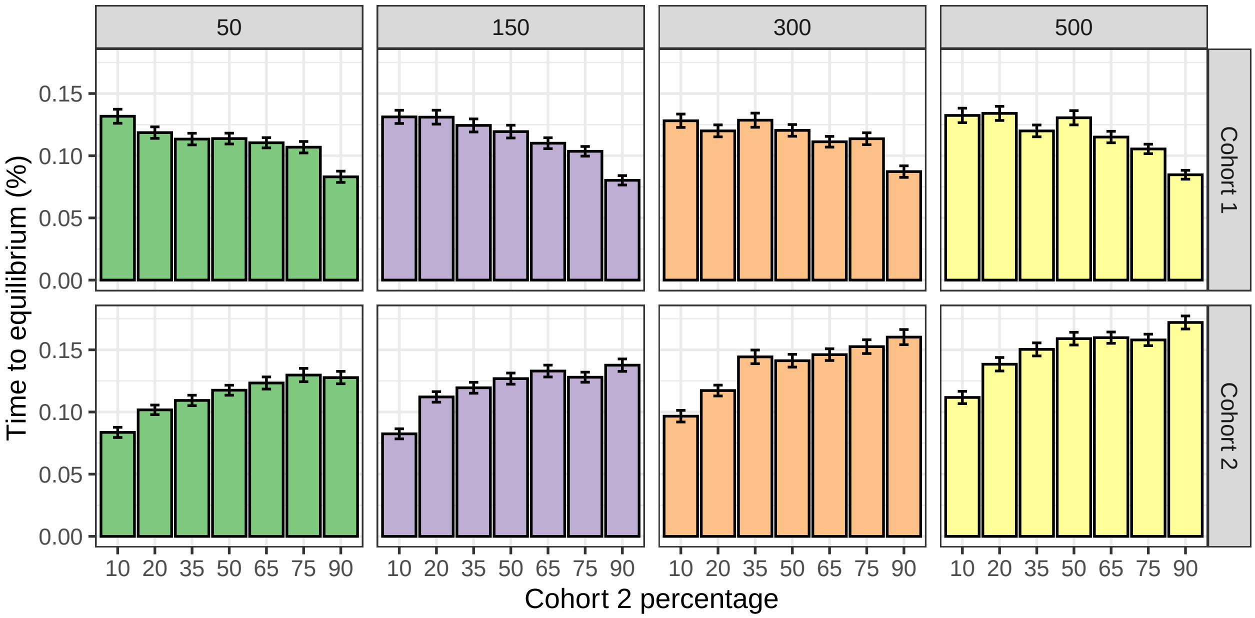

Figure 10: Time to equilibrium for the different age cohorts, faceted by cohort 2 age.

In summary, the older cohort slows down the time to equilibrium, and as the age gap increases, the impact on time to equilibrium increases as well.

Between-cohort communication

For this last set of simulations, gamma was set to decrease with age at a perseverance of 1 and the two age cohorts had an age gap of 200, with the second cohort kept at 50%. The main variation was the willingness to listen to out-group. When this value is 1, nodes will always listen and update their state accordingly regardless of the speaker node's cohort. As this value lowers, the probability of the listener node updating their state decreases as well. For example, a *willingness* value of 0.6 means that there is a 60% chance that a cohort 1 listener node would update their state if the speaker was cohort 2.

As seen in Figure 11, as the willingness decreases, the threshold for full spread of a grammar variant increases. Namely, as willingness decreases, the s-curve becomes more linear, meaning that more popular variants may not spread as much throughout the population. For example, when Grammar 1 is distributed throughout 70% of the nodes, it will spread to 96% of the nodes when willingness is 1.0, but will only spread to 88% when willingness is set to 0.1. In other words, it takes higher initial percentages of Grammar 1 to spread throughout the population.

Figure 12 shows the impact of willingness to listen to out-group on time to equilibrium. For a standard simulation (willingness = 1), the time to equilibrium is n-shaped; low time to equilibrium at extreme values of initial percentage of Grammar 1, and peaks at 50% of Grammar 1. Then, as willingness decreases, time to equilibrium increases across all initial starting percentages of Grammar 1, demonstrating that decreased willingness to listen slows the spread of language through a population.

In summary, (decreased) willingness to listen delays the equilibrium and increases the threshold for a variant to spread throughout the population.

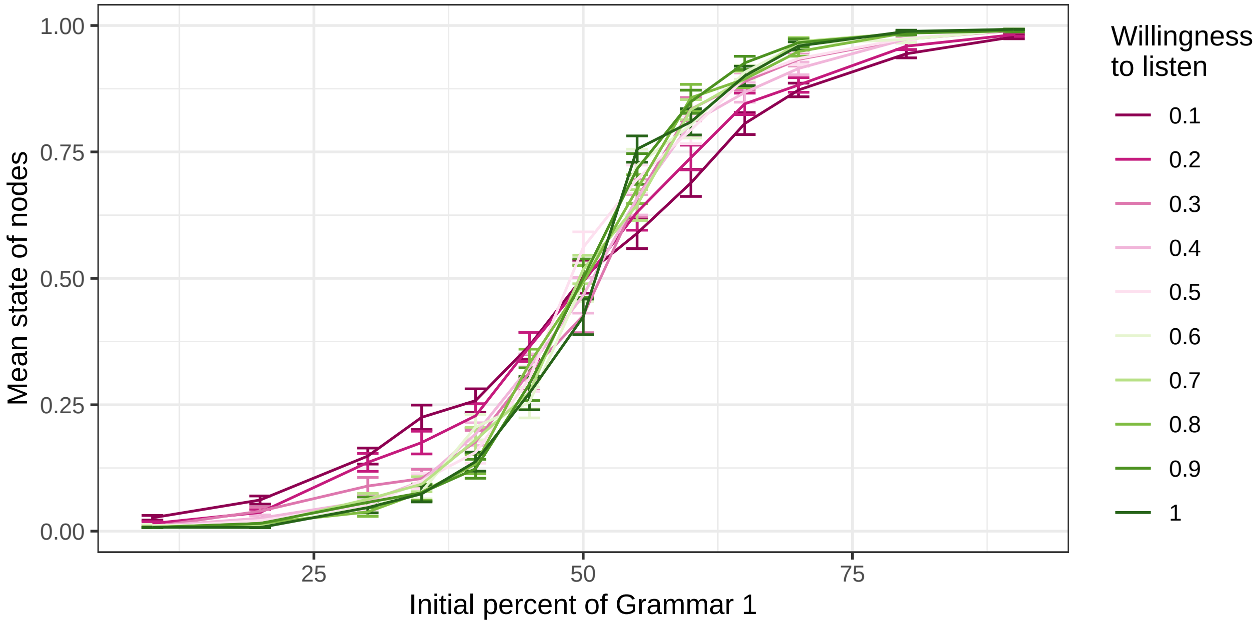

Figure 11: Spread throughout the population based on willingness to listen to the opposite cohort. Color represents the different values of willingness to listen.

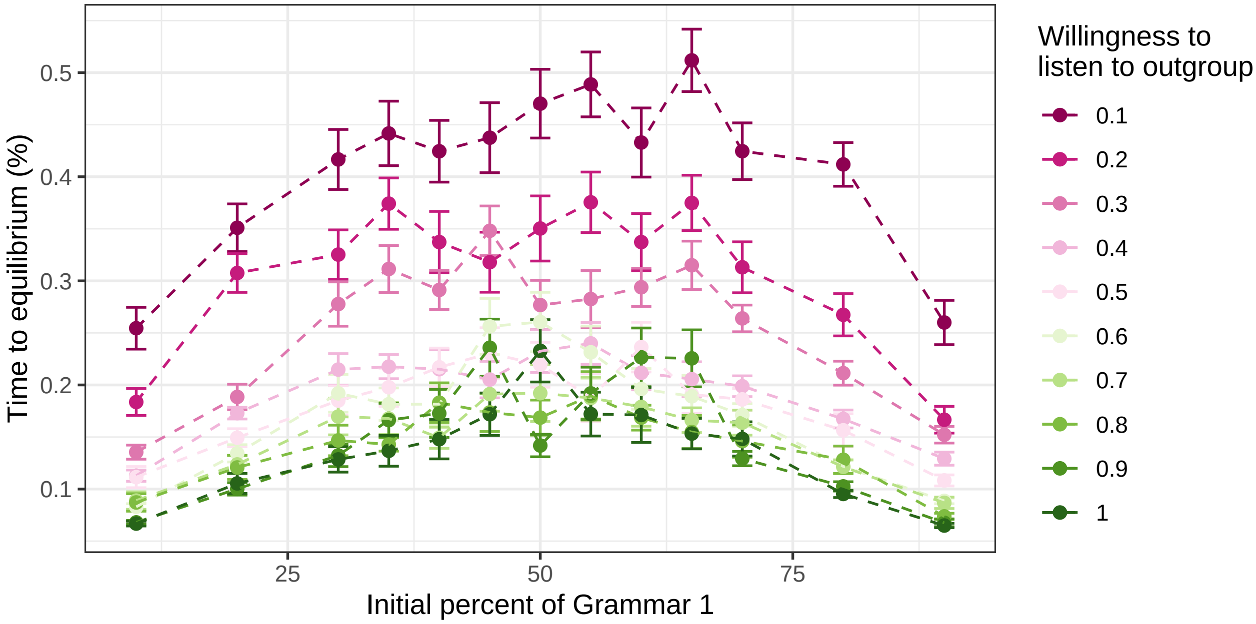

Figure 12: Time to equilibrium based on willingness to listen to the out-group.

Discussion

Overall, these analyses suggest that learning rate mainly impacts time to equilibrium, while group membership factors impact the final state of the system. Variations in learning rate, aging, and perseverance of learning all impacted the time to equilibrium of the system, with little systematic differences in the final state of the system. When group membership factors were added, such that nodes in the network have a chance to ignore out-group members, the relationship between the initial starting percentage of the grammatical variant shifted from a sigmoid to a more linear relationship.

This agent-based model expanded upon previous models of language change by including group variation in learning rates and the addition of age-related dynamics. It also excluded the built-in bias towards a specific grammar variant that previous models included. This model is similar to other agent-based models of language change (Beeksma et al., 2017) or models of opinion dynamics (Castellano et al., 2009; Galesic et al., 2021), but differs with the addition of age-based learning rates.

Extensions

Network structure. One extension would be to implement a small world network structure, where the parameters can be used to vary between fully random and a small world network. I suspect that the emergence of dialect sub-groups (Fig. 2B) is an artifact of the preferential attachment structure, such that you have one well connected node with some other small off-shoot groups. Based on this, there may be a section of the network that is only connected by one node, which will by structure block the grammar from spreading through to the distant group if that one node stops learning. The case for dialect sub-group formation would be stronger if dialect sub-groups still appear when using a small world network.

Cohort-based network structure. In addition to a small world network, another extension would be to implement a connection preference, such that nodes will be more connected to other nodes in the same cohort. More pronounced cohort structure may lead to innovator-reservoir dynamics similar to Turner et al. (n.d.).

Generational mechanisms. In the current model, there is no population shifts; each node is immortal and is never replaced. Future versions could implement generation dynamics and node death, such that nodes reproduce offspring nodes with a similar grammatical variant. Moreover, different types of learning (vertical, oblique, horizontal) could also be integrated, since the current model does not differentiate between them. The *willingness* parameter was aimed at differentiating horizontal and vertical, such that nodes would prefer horizontal to vertical learning.

Learning algorithm. Currently, the learning algorithm is a linear reward-penalty algorithm (Bush & Mosteller, 1951, 1955; Yang et al., 2002), which was one of the defaults from the original model and was tied to the speaking algorithm. One implication of using this algorithm is that when a listener node fails to comprehend an utterance (the grammar it chooses to interpret does not match the utterance), it is punished and the internal state is updated away from the heard grammar. While this may reflect instances in the real world (confusion from misunderstanding as "punishment"), future iterations could change this learning algorithm, such that failure to comprehend is not punished.

Conclusion

In conclusion, this project modified an agent-based model of language change to include parameters for learning rates, aging, and willingness to learn from out-group members. Our main finding was that learning rate impacts the time to equilibrium, where equilibrium is the final state that the system converges to, regardless of whether a grammatical variant fully spread through the population. This flexible definition allowed for inclusion of instances where dialect sub-groups formed in the population. Moreover, if learning rate is based on age, the age gap between groups will modulate the impact of having more older members of the speech community when it comes to language change. Lastly, willingness to listen to out-group members impacts the threshold for and speed of spread.

References

Beeksma, M., Vos, H. de, Claassen, T., Dijkstra, A., & Kemenade, A. van. (2017). *A probabilistic agent-based simulation for community level language change in different scenarios.*

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, *58*(5), 313.

Bush, R. R., & Mosteller, F. (1955). *Stochastic models for learning.*

Bybee, J. (2015). *Language change.* Cambridge University Press.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, *81*(2), 591.

Castro, N., Curley, T., & Hertzog, C. (2021). Category norms with a cross-sectional sample of adults in the united states: Consideration of cohort, age, and historical effects on semantic categories. *Behavior Research Methods*, *53*, 898–917.

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, *53*(8), 1560–1570. https://doi.org/10.1037/dev0000347

"Five Graces Group", Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Galesic, M., Olsson, H., Dalege, J., Does, T. van der, & Stein, D. L. (2021). Integrating social and cognitive aspects of belief dynamics: Towards a unifying framework. *Journal of the Royal Society Interface*, *18*(176), 20200857.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, *331*(6014), 176–182. https://doi.org/10.1126/science.1199644

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(5), 781.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514.

Troutman, C., Clark, B., & Goldrick, M. (2008). Social networks and intraspeaker variation during periods of language change. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 25.

Turner, M. A., Signleton, A. L., Harris, M. J., Harryman, I., Lopez, C. A., Arthur, R. F., Muraida, C., & Jones, J. H. (n.d.). *Minority-group incubators and majority-group reservoirs for climate change adaptation.*

Yang, C. D. et al. (2002). *Knowledge and learning in natural language.* Oxford University Press on Demand.