Social Networks and Intraspeaker Variation During Periods of Language Change

Celina Troutman, Brady Clark, and Matthew Goldrick

1 Introduction

Previous work has revealed general characteristics of language change at both the level of linguistic communities as well as individual speakers. What are the properties of language users such that we can account for these characteristics? To address this question, we built a computational model of a social network of language users. By holding the network structure constant and varying properties of the language users, we found that language change reflects both the structure of social networks and properties of language users. In particular, our results suggest that although language users must be capable of probabilistically accessing multiple grammars, they must prefer to access a single grammar categorically.

1.1 Characteristics of Language Change

To ground our discussion of language change, consider the rise of periphrastic *do* (or *do* support) in English (Ellegård 1953, Kroch 1989, Warner 2004). Prior to about 1400, negative declarative sentences were formed by following a simple finite verb with *not*, as in (1). This was followed by a period of variation from 1400 to 1800 between the older form and the modern form with periphrastic *do*. Importantly, during this time both the older and modern forms were available for a single person, as illustrated in (2).

(1) …whiche he perceiueth not.
(cited in Kroch 1989: 15)
(2) a. I question not your friendship…
(Thomas Otway, "The Cheats of Scapin", 1676/7)
b. She does not deserve it…
(Thomas Otway, "Friendship in Fashion", 1678)
(cited in Warner 2004: 229)

This paper focuses on the following general characteristics of language change, each of which is illustrated by the development of periphrastic *do*:

- *S-shaped curve*: The time course of the change follows an S-shaped curve (Bailey 1973, Kroch 1989): change happened slowly at first, then proceeded very rapidly before slowing down again[1].
- *Intraspeaker variation*: As a new form spreads, speakers do not suddenly jump from always using the older form to always using the new one. Instead, change is gradual, and, as illustrated in (2), there is always a period of intraspeaker variation in which both forms are available to a single speaker (Weinreich et al. 1968).

[1]Note that this characterizes the general trend of language change. For example, in the case of periphrastic *do*, the rate of change varied for different contexts.

- *Categorical norms*: When two syntactic variants are in competition, speakers often move toward categorically using just one of the competing variants (Kroch 1994). For example, in present day English, speakers categorically use periphrastic *do* in negative declaratives.
- *Multistability*: Language change can have multiple stable outcomes (Clark et al. in press). For instance, in the history of English, initially rare periphrastic *do* spread through the entire speech community, but this was not the only possible outcome. Under different circumstances, periphrastic *do* could have been used for only a short time before fading away. Reverse movements (A > A/B > A rather than A > A/B > B) are always possible in language change (Fischer 2007: 192).
- *Threshold problem*: Initially rare variants, such as periphrastic *do*, manage to spread to entire speech communities. However, this is counterintuitive because learners should adapt their speech to match their environment. If the majority of the population is still using the older form, a learner should adopt that form as well. Learners should never use *more* of the minority form than the rest of the population. Nettle (1999) has referred to this issue as the threshold problem: how can an initially rare variant (e.g. periphrastic *do*) spread through a speech community (Sapir 1921)?

1.2 Previous Work

To understand the conditions necessary for language change to occur, analytical (Watts 2002) and simulation (Nettle 1999; Kirby 1999) studies have explored the conditions under which an initially rare variant can spread through an entire population (i.e., conditions for solving the threshold problem). (Note that Watts does not focus on linguistic change specifically, but on the spread of innovations through a network.) These models share two key assumptions about the nature of language users. In all three models, individuals have discrete grammars, meaning they have access to only one grammar at a time. (This was represented in Watts' (2002) model by assigning people to one of two discrete states—they have either adopted or not adopted the innovation.) Second, these models incorporate some kind of bias in favor of the initially rare variant, either explicitly or implicitly. In some models, learners are more likely to acquire the initially rare variant (e.g., because it is associated with prestigious speakers—Nettle 1999; or it is functionally preferred—Kirby 1999). Others incorporate the additional assumption that once learners acquire the initially rare variant, they never return to using the older form (Watts 2002).

In the next section, we discuss a model of language change that incorporates the assumptions of discrete grammars and bias for the initially rare variant. We demonstrate that although this model captures most of the characteristics of language change discussed above, it cannot capture intraspeaker variation. In Section 3, we show that simply incorporating probabilistic grammars into the discrete model fails to account for multistability. Finally, in Section 4, we present a probabilistic model that captures all of the key characteristics of language change.

2 Discrete Model of Language Change

To simulate language change in a speech community, we used NetLogo, a multi-agent programmable modeling environment[2]. Our computational model has three main components to the model: the language users, the social network structure, and the learning algorithm.

---

[2] http://ccl.northwestern.edu/netlogo/

2.1 Language Users

In this model, language users can have only one of two types of grammars. We refer to these as the *+DO grammar* and the *-DO grammar*. Note that the model is not intended as a complete theory of development of periphrastic *do*. There are many complexities associated with that change (see, e.g., Kroch 1989 and Warner 2004). Our model is simply intended to capture the competition between forms of any sort (e.g. negative declaratives with and without periphrastic *do*) during periods of language change.

In the discrete model, speakers produce utterances in accord with a single grammatical option. For example, speakers always produce sentences with *do* support (e.g. *she does not deserve it*), or without *do* support (e.g. *she deserves not it*), but no single speaker produces both.

2.2 Social Network

Language users are connected to each other in a social network. Networks are constructed through the process of "preferential attachment" in which individuals enter the network one by one, and prefer to connect to those language users who already have many connections (Barabási and Albert 1999). This leads to the emergence of a few "hubs", or language users who are very well connected; most other language users have very few connections.

Figure 1 shows a miniature version of the type of social network used here. Circles represent language users, and lines represent the connections between them. Language users only interact with those they are directly connected to. Each circle's color represents the individual's grammar. Black circles represent speakers who never use periphrastic *do*, and white circles represent speakers who always use periphrastic *do*. Note in the middle of the network there is a hub speaker connected to seven others. If another speaker were to enter this network, they would be likely to connect to the hub speaker. However, it is also possible to connect to less-popular members of the network (leading to the occasional creation of side branches).

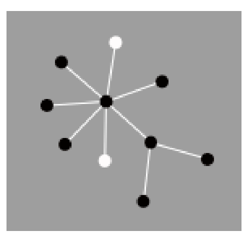


Figure 1. Miniature social network

We chose to model communities with this type of network structure because a number of networks tend to have a few well-connected items and many less-connected ones (Barabási 2003). For example, personal relationships, the Internet, and networks of academic paper citations all display this characteristic structure. Additionally, our network falls into a larger class of "scale-free" networks which share a number of mathematical properties (Barabási 2003). This suggests the results discussed here may be generalized to other network structures; they are not necessarily limited to those generated through the process of preferential attachment.

2.3 Learning Algorithm

Language users interact with each other based on who they are connected to in the network. At each iteration, everyone *speaks* by passing an utterance either with or without *do* to their neighbors in the network. Individuals then *listen* to their neighbors by changing their grammars based on what they received as input from the speakers. The order that individuals listen is randomized for each iteration, and each individual updates its grammar immediately after listening. Following previous models discussed above, speakers are biased towards adopting the initially rare variant. Specifically, learners adopt the +DO grammar if they hear utterances with *do* support from at least 30% of their neighbors. Otherwise they adopt the -DO grammar.

2.4 Results

We generated networks consisting of 40 people, running each network for 12 iterations of speaking and listening. We ran a total of 1000 networks, generating a new instance of the same network type for each run. This insured that the results would not be an artifact of any particular network structure, but would instead reflect the general behavior of scale-free preferential attachment networks. For each run, individuals' grammars were initialized so that 25% began with the +DO grammar and the remaining 75% were initialized with the -DO grammar.

Figures 2a and 2b demonstrate the results of two typical runs. The x-axes show the number of iterations and the y-axes show the proportion of language users that have the +DO grammar. This model was able to capture four out of the five characteristics of language change:

- *S-shaped curve*: Both Fig. 2a and Fig. 2b resemble the S-curve, the time-course of change observed by Kroch (1989) and others.
- *Categorical norms*: At the end of each simulation run in Fig. 2a and 2b, language users converged on the same grammar, +DO or –DO.
- *Multistability*: While the speech community converged on the +DO grammar in Fig. 2a, it converged on -DO in Fig. 2b.
- *Threshold problem*: In Fig. 2a, the initially rare +DO grammar spread to everyone in the network.

However, by design, language users do not exhibit intraspeaker variation, since they have access to only one grammar at a time. We therefore modified the model to incorporate the assumption that linguistic knowledge is probabilistic, rather than discrete.

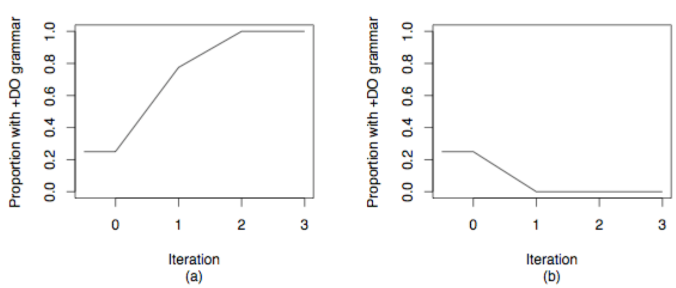


Figure 2. Proportion of +DO speakers vs. iteration for the discrete model

3 Probabilistic Model of Language Change

In this model, the social network structure remained the same as described in Section 2.2, but the representation of the language users and their learning algorithm was changed to accommodate probabilistic grammars.

3.1 Language Users

In this model, individual language users can access both grammars. Each grammar is associated with a weight, which determines the language user's probability of accessing that grammar. However, because there are only two grammars in competition, the weights in our model are represented with a single value—the weight of the +DO grammar. Speakers still produce utterances in accord with the grammar accessed, but individuals now have a *probability* of producing sentences with or without *do* support. This allows us to capture intraspeaker variation during language change.

3.2 Learning Algorithm

At each iteration, language users speak and their immediate neighbors listen and update their grammars based on what was heard. Speaking involves choosing a grammar based on its weight. As before, individuals have a bias in favor of choosing the +DO grammar. This bias is implemented by increasing each speaker's probability of using *do* by a small amount (weight of +DO grammar * 0.5) at every speaking event. Figure 3 shows the relationship between the weight of the +DO grammar and an individual's probability of selecting that grammar. For instance, if the weight is 0.2, a speaker will select that grammar with a probability of 0.3. If the weight is greater than approximately 0.67, the probability of selecting the +DO grammar will always be 1.0.

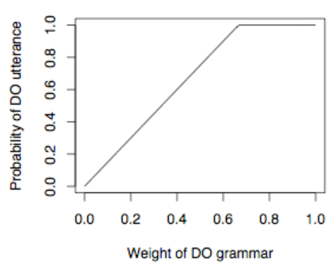


Figure 3. Probability of *do* utterance vs. weight of +DO grammar

Once an individual speaks, its neighbors in the network listen and update their grammar weights according to the linear reward-penalty algorithm (Bush and Mosteller 1951, 1958, Yang 2002). In this algorithm, a learner probabilistically selects a grammar to analyze an utterance spoken by its neighbor (where the probability of selecting a grammar is equal to its weight). If the selected grammar can successfully analyze the utterance, the grammar is rewarded by increasing its weight. Otherwise, the grammar is penalized by decreasing its weight (see Yang 2002 and Clark et al, in press for details on the implementation of this algorithm). In short, if an individual hears an utterance with *do* support, the individual's weight of the +DO grammar is increased, and they will be more likely to access the +DO grammar in the next iteration. Similarly, hearing an utterance without *do* support increases the likelihood of accessing the -DO grammar in the next iteration.

3.3 Results

We generated 1000 networks consisting of 40 individuals each, running each network for 1000 iterations. Like the discrete model in Section 2, these networks were initialized so that 25% of language users began with the weight of the +DO grammar equal to 1, meaning they could only access that grammar. The remaining 75% were initialized to only have access to the -DO grammar. Figures 4 and 5 represent two typical runs of this model. The results show that this model can capture four out of the five characteristics of language change discussed above:

- *S-shaped curve*: The time course of change always followed an S-shaped curve.
- *Intraspeaker variation*: Individuals produced utterances both with and without *do* support. This is illustrated in Fig. 4, which shows how the distribution of individuals' weights for the +DO grammar changed over time. The first column represents the initial state of the network, in which 25% of people have a weight of 1 for the +DO grammar, and the rest have weight of 0. The second column shows that after 100 iterations, people have a range of intermediate weights, indicating the presence of intraspeaker variation.
- *Categorical norms*: At the end of the run in Fig. 5, the mean weight of the +DO grammar is 1. All language users therefore categorically produce one form (e.g. negative declaratives with *do*).
- *Threshold problem*: The community eventually converged on grammars that categorically produced the initially rare +DO form.

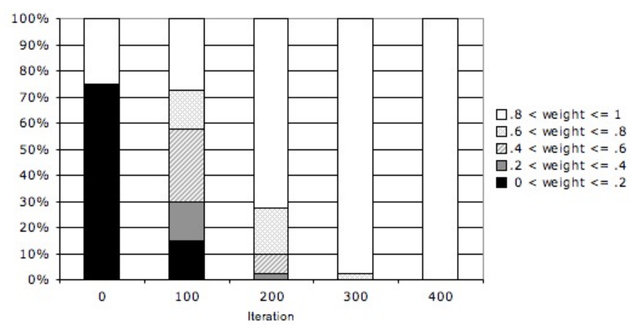


Figure 4. Proportion of speakers with different grammar weights over time

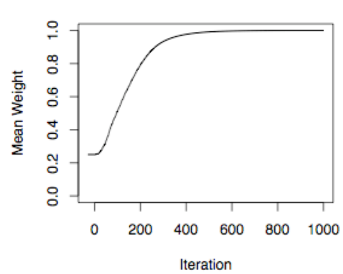


Figure 5. Mean weight of +DO grammar vs. time for probabilistic model

Recall the discrete model in Section 2 incorrectly rules out intraspeaker variation during language change. The probabilistic model explored in this section captures intraspeaker variation but wrongly rules out multistability. In all 1000 runs, individuals converged on categorically using the *favored* variant only. One might think that if the bias for *do* support was lowered that multistability would emerge. However, varying the amount of bias for the +DO grammar only affected the rate of change, never its direction. In the next section, we present a model that captures all five characteristics of language change discussed in Section 1.1.

4 Probabilistic Model with Preference for Discrete Grammars

The model discussed in this section shares the social network structure of the previous models and the probabilistic grammars of the model in Section 3. The learning algorithm in Section 3 was modified to incorporate a soft preference for discrete grammars. This preference is motivated by research suggesting that even when multiple options are available in the linguistic environment, individuals prefer to use only a single grammatical option. For instance, Kroch (1994) has proposed that when syntactic forms are in competition, there is pressure over time for one to win out due to a "blocking effect"[3]. Additionally, work by Elissa Newport and colleagues (e.g. Singleton and Newport 2005, Hudson Kam and Newport 2005) has shown that language learners have a dispreference for acquiring stochastic patterns.

To implement this preference for discrete grammars, each speaker's weighting of their grammatical options was skewed towards extreme values. Figure 6 shows the relationship between the weight of the +DO grammar and the probability of uttering the *do* variant for this model.

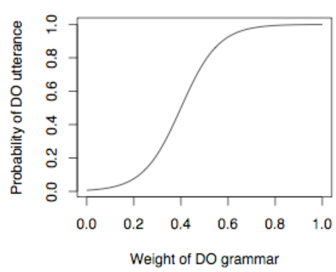


Figure 6. Probability of DO utterance vs. weight of +DO grammar

For example, if the weight of +DO grammar is 0.6, the probability of uttering *do* will be pushed even higher to 0.9. However, if the weight of +DO grammar is 0.2, the probability will be reduced to 0.1.

In addition to a preference for discrete grammars, this model includes the bias for *do* support that was part of the models discussed in Sections 2 and 3. This bias shifts the inflection point of the curve in Fig. 6 slightly to the left. For example, for a grammar weight of 0.50, the probability of uttering *do* is about 0.78 (see Clark et al., in press, for implementation details)

[3] This effect is analogous to the blocking effect in morphology, which acts to prevent the coexistence of forms that are equivalent in meaning.

4.1 Results

The procedure for generating and running networks was identical to the procedure for the probabilistic model in Section 3. Figure 7 demonstrates how change proceeded for two runs of this model. Our results indicate that unlike the previous two models, this model could capture all five characteristics of language change:

- *S-shaped curve*: The time course of change always followed an S-shaped curve.
- Intraspeaker variation: Individuals produced utterances both with and without *do* support.
- *Categorical norms*: By the end of the run in Fig. 7a, the mean weight for the +DO grammar was nearly 1, while by the end of the run in Fig. 7b, the mean weight was nearly 0. In both cases, the language users moved toward categorically using the same form.
- *Multistability*: While the +DO grammar took hold in the run in Fig. 8a, the -DO grammar remained dominant in Fig. 7b.
- *Threshold problem*: In Fig. 7a, the entire speech community eventually converged on grammars that categorically produced the initially rare +DO form.

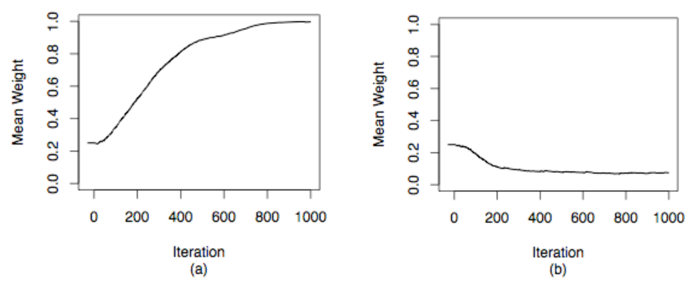


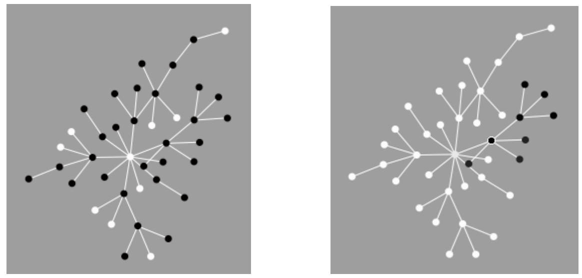
Figure 7. Mean weight vs. iteration for discrete model

4.2 Emergence of Dialect Subgroups

So far we have discussed outcomes of the model in which the entire population converged on a single grammar. However, in some simulation runs, subparts of the network converged on different linguistic models.[4] Figures 8a and 8b show a network before and after a run of 1000 iterations. Over time, the initially rare +DO grammar (represented by white circles) spread through the majority of the network, but one subgroup (the black circles in Fig. 8b) resisted the change. Importantly, language users in each group did converge to a categorical norm—they ended up with a weight of approximately 0 or 1—but this norm was not shared by all speakers in the network. The only exception is a language user who is connected to more than one group (e.g. the black circle with a white border in Fig. 8b). Since this speaker continues to receive both variants as input, its weight remains at an intermediate value. This situation illustrated in Fig. 8 may be viewed as the emergence of dialect subgroups.[5]

[4]It was also possible for subgroups to emerge in the discrete model of Section 2.

[5]To test the extent to which language users formed separate dialect groups, we employed Newman and Girvan's (2004) measure of the *modularity* of a network. We simulated an additional 24 networks (following Section 4) and calculated the modularity of the networks before and after each run. A paired t-test showed that the final states (mean $Q = 0.13$) were significantly more modular than the initial states (mean $Q = 0.0$; $t(23) = 3.8$, $p < 0.001$).



(a) (b)

Figure 8. Initial state (a) and final state (b) of a network

5 Discussion

Our goal was to develop a computational model that captures the five key characteristics of language change discussed in the Introduction. We investigated what properties language users must have in order to account for these key features. The discrete model fails to capture a key property of language change (intraspeaker variation), but simply incorporating probabilistic grammars into the discrete model fails to account for multistability. However, when learners have probabilistic grammars combined with a preference for having discrete grammars, all five characteristics of language change can be captured.

Our results accord with those of Clark et al. (in press), who used a model of language change to explain the emergence of typological word-order correlations. They argued for similar constraints on language users, such as a soft preference for discrete grammars and a bias for the typologically preferred variant. Additionally, our results are consistent with Pearl and Weinberg (2007) who demonstrated that successful modeling of historical language change data from Old English requires that there be a "filter" on a probabilistic learner's input. This filter restricts the learner's attention to a particular subset of their input leading to effects similar to those of bias in our model (i.e., causing the learner's grammar state to be a non-veridical reflection of the total set of input data).

Both Clark et al. and Pearl and Weinberg's models examined unstructured populations, simulating interactions in a random network. An advantage of our model is the incorporation of a more realistic social network, limiting language users' input to a small number of individuals rather than the entire population. This allowed for the emergence of dialect groups in our model. In contrast, in the dynamic random networks of Clark et al. and Pearl and Weinberg, the entire population always converged to a single grammar. Further work is needed to better understand when exactly subgroups can arise in our model.

5.1 The Role of the Bias

In designing our model, we followed previous work that included a bias for the initially rare variant. For instance, in Nettle's (1999) model, if there was no preference to acquire the variant associated with prestigious speakers, the threshold problem could not be solved. Pearl and Weinberg (2007) also found that without a bias (or filter) on the learner's input, their model's output failed to match the observed historical data. Additional exploration of our own simulations revealed similar findings[6]. Although such results suggest that a bias is a critical component of models of language change, it remains unclear what source(s) underlie these effects. Some have attributed biases to social structure (e.g., Nettle 1999) while others have attributed them to properties of perception/production processes (e.g., Kirby 1999). Future work should examine the relative ability of these contrasting perspectives to account for the properties of language change.

5.2 Future Work

Our simulations focused on cases where a small percentage of a population initially uses one grammar (G1) categorically, and the rest uses G2 categorically. This could represent the starting state for a language contact scenario. However, in the case of *do* support, speakers initially used periphrastic *do* at less than categorical rates (Kroch 1989). (This scenario is common to many documented cases of language change.) To develop a more accurate model of this type of change, a small percentage could initially use G1 variably, and the rest use G2 categorically. The framework developed in this paper would enable us to easily explore this condition in future work.

References

Bailey, Charles-James. 1973. *Variation and Linguistic Theory*. Washington: Center for Applied Linguistics.

Barabási, Albert-Laszlo. 2003. *Linked: How Everything is Connected to Everything Else and What it Means*. Plume.

Barabási, Albert-Laszlo, and Reka Albert. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Bush, Robert R. and Frederick Mosteller. 1951. A mathematical model for simple learning. *Psychological Review* 68:313–323.

Bush, Robert R. and Frederick Mosteller. 1958. *Stochastic models for learning*. New York: Wiley.

Clark, Brady, Matthew Goldrick, and Kenneth Konopka. in press. Language change as a source of word order correlations. In *Language Evolution: Cognitive and Cultural Factors*, ed. R. Eckardt, G. Jäger, and T. Veenstra. Berlin: Mouton de Gruyter.

Ellegård, Alvar. 1953. *The Auxiliary Do: The Establishment and Regulation of its Use in English*. Gothenburg Studies in English. Stockholm: Almqvist and Wiksell.

Fischer, Olga. 2007. *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: Oxford University Press.

Hudson Kam, Carla, and Elissa Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1:151–195.

Kirby, Simon. 1999. *Function, Selection and Innateness: the Emergence of Language Universals*. Oxford: Oxford University Press.

[6]To examine if a bias was necessary to solve the threshold problem, we simulated a model with a preference for discrete grammars without a bias towards the initially rare DO support variant. The network size, structure and initial grammar distributions followed the simulations above. The final mean grammar weight exceeded 0.4 in only 5/100,000 simulations (maximum: 0.54). This suggests that without a bias an initially rare variant cannot come to dominate the entire population.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

Kroch, Anthony. 1994. Morphosyntactic variation. In *Proceedings of the thirtieth annual meeting of the Chicago Linguistics Society*, ed. K. Beals, Vol. 2. 180–201.

Nettle, Daniel. 1999. Using Social Impact Theory to simulate language change. *Lingua* 108:95–117.

Newman, M.E.J. and M. Girvan. 2004. Finding and evaluating community structure in networks. *Physical Review E*. Vol 69. Part 2. 026113.

Pearl, Lisa, and Amy Weinberg. 2007. Input filtering in syntactic acquisition: Answers from language change modeling. *Language Learning and Development*, 3(1), 43–72.

Sapir, Edward. 1921. *Language: an introduction to the study of speech*. New York: Harcourt, Brace, and World.

Singleton, Jenny L., and Elissa Newport. 2004. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology* 49:370–407.

Warner, Anthony. 2004. What drove DO? *Amsterdam Studies in the Theory and History of Linguistic Science*, Series 4. Vol. 251. 229–242.

Watts, Duncan. 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* USA 99, 5766–5771.

Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics: A Symposium*, ed. W. P. Lehmann, 95–195. Austin: University of Texas Press.

Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Department of Linguistics
Northwestern University
Evanston, IL 60208
c-troutman@northwestern.edu
bzack@northwestern.edu
matt-goldrick@northwestern.edu