Language and Linguistics Compass 2/2 (2008): 289–307, 10.1111/j.1749-818x.2008.00054.x

Computational Approaches to the Study of Language Change

Adam Baker*
University of Arizona

Abstract

The article reviews computational studies of language change. Computer models of change are helpful because of the complexity of the behavior involved: an entire population of complex, interacting agents must be accounted for. Computational studies frequently bring to light hidden implications of theories, which make them relevant to the theoretical development of both acquisition and change. Studies of language change have focused on discovering mathematical properties of dynamical systems, or on simulating populations of speakers that interact with one another and change their internal states as a result. Models of lexical (including phonological) and syntactic change are considered. Computational models of change have proved useful tools for testing theories of language change, and will prove more useful as the field matures to include more systematic studies of the effects of varying model parameters in complex simulations.

This article reviews research developments in the computational studies of language change.[1] Computational models are important to the investigation of language for several reasons. First, they provide quantitative results. Assuming that a quantitative measure of success can be determined, different models of the same phenomenon can be compared to one another objectively. Second, models require precise theories. Implementing the assumptions of a theory in computer code permits no sloppiness in definitions or ideas. Thus, even the construction of a model can uncover aspects of a theory that need to be made more explicit.

The fundamental purpose for developing a model of a phenomenon is to increase our understanding of it. Nigg (1994) notes an apparent contradiction here: while models are expected to broaden knowledge, they consist solely of the knowledge that the modeler has encoded.[2] He explains, however, that the contradiction is based on a misunderstanding of models: the purpose of a model is not to generate novel behavior, but rather to better understand relationships between variables in the system. Put differently, the purpose of a model is to increase our understanding of a particular theory. Encoding one's assumptions into a computational model is an excellent way to discover otherwise hidden implications of one's assumptions. Several examples of such discoveries will be presented in this article.

Computational models are particularly appropriate tools for studying language change. Observing language change is difficult, and the majority of historical changes have occurred in cultures without written records. Computational studies are also appropriate because they enable researchers to observe the 'behavior' of many interacting individual models of speakers. Faced with the enormously difficult problem of accounting for humans' linguistic ability, linguists have posited that speakers perform very complex operations (e.g. Ohala 1993, who provides evidence that speakers are able to perceptually factor out the effects of other speakers' co-articulation – presumably leveraging knowledge of their own vocal tract – in order not only to determine the intended message, but also the speaker's intended pronunciation). A study of language change can predict how these complex abilities interact in a population of speakers, a task made considerably easier with computer simulations. No model of change approaches the level of algorithmic complexity that must exist in the language faculty, but importantly, even when very simple assumptions are made about individual speaker's abilities, unexpected results can arise in the behavior of the population. Computational studies of change have also proved relevant to theories of acquisition, since, under the assumption that imperfect acquisition drives change, a model of change can be used to validate a theory of acquisition by comparing predicted to attested changes in languages.

The general approach studying language change computationally is to model a population of individual speakers that change their internal parameters in response to their experience. Models vary according to the number of speakers, their organization in a social network, and the behavior of speakers. Most frequently the state of the model is described as the proportion of speakers who have adopted a certain variant (in the case of lexical change), or a certain grammar (in the case of syntactic change). In describing these models below, specific assumptions and parameters have been noted only where they seemed important to the behavior of the model. The reader is referred to each article for a full description of their respective details. To organize the discussion, models of lexical change are considered first, followed by models of syntactic change.

1 Models of Lexical Change

Models of lexical change focus either on the prevalence of two competing lexical variants, or of the diffusion of a feature through the lexicon. The models are generally sufficiently abstract that the precise feature being modeled need not be specified. Competing variants could either be lexical items (e.g. *soda* and *pop*), or different pronunciations of the same word (e.g. *libary* and *library*).

Nettle (1999a) presents a computation model of language change using Social Impact Theory (Latané, 1981). He considers the general case where two speech variants, *p* and *q*, are competing within a speech community. He models a population of 400 speakers that move through a lifecycle stage at each time increment, which represents aging. At each increment, individuals interact with others of their age group (representing peer interactions), and with individuals that are immediately older and younger (representing family interactions). In the first two life stages, individuals adopt either *p* or *q*, based on that individual's estimation of which variant has the greater social impact, a concept taken directly from Social Impact Theory. (To simulate imperfect learning, with 5% probability the learner adopts the variant with the lesser social impact.) The values i_p and i_q respectively describe the social impact of *p* and *q*. i_p is calcuated with the following equation (with i_q being calculated in an analogous way).

$$i_p = b_p N_p^a \sum_{j=1}^{N_p} \frac{s_j d_j^{-2}}{N_p} \tag{1}$$

Here, b_p is a possible phonetic or cognitive bias that favors or disfavors the *p* variant (for $b_p>1$ and $b_p<1$, respectively). N_p is the number of individuals that have *p*, which is raised to the a^{th} power. This means that the impact of the *p* variant increases as more people adopt it. This relationship can easily be imagined to be non-linear, and the term *a* allows for this kind of relationship. The summation is a weighted average describing the average social standing (*s*) of individuals who have adopted *p*. d_j is the distance between the individual and the j^{th} individual who has adopted *p*. Note that scaling s_j by d_j^{-2} means that the influence that one individual has over another decreases in inverse proportion to the square of the social distance between them.

In Nettle's simulations, every speaker starts with *p*. As the simulation progresses, the 5% error rate in learning guarantees that *q* is learned occasionally; thereafter, other individuals have the potential to adopt *q* through social diffusion. From multiple simulations with $a = 1$ and without a functional bias or variation in social status, Nettle found that the *q* variant never takes hold in the population. But if *a* is set to 0.5 or less, which diminishes the effect of population size in calculating i_p and i_q, then about half the population adopts *q*.[3]

In simulations where different individuals had different social status, Nettle found that variation in social status affects simulations only if the difference in social status is very large, such that the range of social status is 10,000-fold. In this case, achieving language change was essentially a matter of having a hyper-influential person adopt the *q* variant, which then spread through the entire community.

When a bias was introduced in favor of or against *q* (respectively, $b_q = 1.1$ and $b_q = 0.9$), the model becomes deterministic: a bias against ensured that *q* disappeared, and a bias in favor guaranteed that it prevailed. This was only the case, however, when a status differential was maintained: if the status differentials were removed, then even a bias in favor of 2 was insufficient to induce change.

Nettle (1999b) used the model to answer a few specific questions about language change, following the findings of Nichols (1990, 1992), who explored the variation and causes of linguistic diversity in different geographical areas. Simulations are appropriate to this kind of study, since direct observation of historical diversification is impossible. He found that language change proceeds more slowly with larger populations of individuals (varying from 100 to 500). In larger populations, the inertia of the status quo is apparently greater. He also created a model of the Norman invasion, by adding a few 'foreign' individuals that have the *q* variant and very high social status. The effect of varying the size of the population was similar, in that smaller communities were more susceptible to accepting borrowed variants. He finally found that small communities (to a much greater extent than large ones) can better maintain marked structures. A sufficiently small population can maintain a variant with a disfavoring bias as great as 0.8, for considerably longer than a larger population.

Nettle's model was adopted by Culicover et al. (2003), who applied it to find an explanation for gaps in language typologies. Generative accounts of language generally attribute such gaps to innate properties of the language faculty, but Culicover et al. suspected that such gaps might arise accidentally, as a result of historical change. The simulations they performed indicate that such gaps in a typology may arise accidentally when particular features become geographically isolated: the rarer they are, the less likely they are to overlap with other features. Indeed, in their simulations, it was common for only a proper subset of possible languages to be attested, although no relationship between features of the language was encoded into the model.

Baker (2008) created models similar to Nettle's to investigate how different assumptions about speakers influence the rate of sound change, which was assumed to be uniform across the lexicon. He provides a computational model that follows the Neogrammarians (and more recent computational models, for example, Pierrehumbert 2002) in attributing sound change to the phonologization of co-articulation. Multiple simulations under this assumption invariably produce sound change, which is counterfactual; this was a computational recapitulation of a point originally made by Weinreich et al. (1968). Baker also performed simulations to determine the implications of different assumptions about how speakers decided to adopt a new variant. He found that sigmoidal progression of change obtained when a speaker's probability of adopting the change was proportional to the prevalence of the change in the speech community.[4] He also notes that modeling a change that requires a long time to complete requires that the ability to acquire a new variant decreases with age (either gradually or abruptly); otherwise changes propagate through the community very quickly. Thus, the presence or absence of a critical period, and the way that plasticity decreases with time, is important to understanding the time course of language change.

Wang et al. (2004), building on the work of Shen (1997), analyze the gradual diffusion of a sound change through the lexicon. Their model follows those of Niyogi and Berwick (see discussion below) in analyzing sound changes in terms of dynamic systems, but is unlike Niyogi and Berwick's work in that they do not restrict themselves to a small number of independent parameters, but instead consider the interaction of many words in the lexicon.

They represent the prevalence of a changed word variant c and the prevalence of an unchanged word variant u as:

$$c(t + \partial t) = c(t) + \alpha c(t)u(t)\partial t \tag{2}$$

$$u(t + \partial t) = u(t) - \alpha c(t)u(t)\partial t \tag{3}$$

In these equations, α represents a phonetic or social bias toward the adoption of the changed form over a simulation increment. As a model for a single word, the system above propagates through the population in a sigmoidal fashion. These results are broadened to include the multiword case below, where c_i, for instance, represents the prevalence of the changed form for the i^{th} word in the simulated lexicon.

$$c_i(t + \partial t) = c_i(t) + u(t)\sum_{j=i}^{n} \alpha_{ij} c_j(t)\partial t \tag{4}$$

$$u_i(t + \partial t) = u_i(t) + u(t)\sum_{j=i}^{n} \alpha_{ij} c_j(t)\partial t \tag{5}$$

Here, α_{ij} takes on a different interpretation. For $i = j$, it represents the bias toward the changed form (as with α above). For $i \neq j$, the term refers to the influence of the j^{th} word on the i^{th} word. It represents the influence that one lexical item has over another (say, in an analogical sense).

In simulations, appropriate values of a_{ij} enable lexical diffusion: the tendency for one lexical item to change can exert pressure on other lexical items. Wang et al. do not provide an exhaustive analysis of the influence of varying different model parameters. They do note, however, that the model has some attractive properties. It captures the generalization, for instance, that words that begin participating in a change relatively late, change at a faster rate than those that started earlier. Wang et al.'s model provides a helpful, and fairly simple, mathematical formalism for representing the diffusion of changes through the lexicons of a population. Clearly, much rests on the values that are chosen for α or α_{ij}, which represent the phonetic biases and cohesiveness of different lexical entries. Since the parameters are readily interpretable, however, the model provides a test for hypotheses about their respective values. One can imagine constructing a model of the changes in a particular lexicon by finding appropriate values of a_{ij}, and using this to objectively determine how classes of words affect one another.

Dras and Harrison (2002) and Harrison et al. (2002) report on simulations of the rise of backness harmony in the Turkish lexicon. They modeled a population of Turkic speakers that interacted with one another. Their simulation began with a 1,000-word lexicon, with 50% of the words harmonic with respect to backness. When encountering a word, each speaker would either harmonize the word or disharmonize the word with some probability, that is, changing the vowels of the lexical entry of the word. In the simulations reported, the authors left the probability of disharmonizing a word at zero, and modeled cases where words were harmonized with non-zero probability. In these simulations, harmony proceeded rapidly and linearly. Aiming at a sigmoidal growth curve, the authors modified this probability so that the propensity to harmonize a word was proportional to the number of harmonic entries in one's lexicon. In this schema, more sigmoidal curves arose. Although the model accounts for the data, their treatment of speakers' phonetic tendencies is a bit heavy handed. Speakers in their model have a tendency to induce harmony, but there is no account of where that tendency comes from (particularly if only half the words of the lexicon are harmonic to begin with), or of why such a tendency should not be present in all humans, which would lead to backness harmony in all languages.

de Boer (1999, 2002) has conducted simulations in which populations of agents play 'language games'. An agent produces a vowel, has another speaker guess the vowel, and corrects the other speaker if necessary. de Boer originally applied the models in studying broad cross-linguistic tendencies of vowel systems. The model has also been applied to the study of language change, however. de Boer (2003) used the model to study the long-term effects of phonetic vowel reduction. In the production stage of the model, he added a small phonetic bias toward reduction. He found that, over time, the cumulative effect of this reduction lead vowel systems to collapse into a single central vowel. Since this is demonstrably not the outcome of vowel reduction in human language, de Boer investigated possible mechanisms that could compensate for the tendency. Two such compensatory mechanisms were tested: first, the proposal that infants assume reduced speech and automatically compensate by perceiving a more expanded vowel space; second, the idea that infant-directed speech (which has a less reduced vowel space, Kuhl et al. 1997) protects children from learning reduced vowels.[5] In simulations of five-vowel systems, either compensatory mechanism was sufficient to preserve the vowel system. This was not true in simulations of seven-vowel systems, however; in this case neither mechanism could prevent reduction of the vowel system. If both mechanisms were combined, however, so that infant directed speech was reduced in proportion to what the infant compensated for, reduction slowed considerably. de Boer does not report on simulations with more than seven vowels.

Pierrehumbert (2002) models phonetically gradual sound change using an exemplar model of the lexicon. In an exemplar model, phonetic representations are fine-grained, and categories are defined by clouds of exemplars in phonetic parameter space. Pierrehumbert assumes that the speaker's production is based on a single exemplar drawn from cloud, and that the speaker's productions are categorized by the speaker and added to the exemplar cloud. In the simulations reported, there is one speaker that talks to itself. She finds that over time, the variance of a speaker's categorizations increases with time. That is, noise generated during production accumulates, leading to a broadening of the category. The effect is mediated somewhat if productions are based instead on the average of a small number of randomly selected exemplars. When phonetic bias is added, the mean value of the category shifts over time in the direction of the bias.

Wedel (2006) also studies language change with exemplar models, considering specifically how they enable linguists to understand language change phenomena within the analytical framework of Darwinian evolution. The assumptions of his exemplar model are similar to those of Pierrehumbert (2002). He notes several factors that limit the influence of noise in determining the shape of exemplar clouds, drawing from evolutionary biology. One mechanism by which noise is reduced is clipping of lines of descent. An exemplar's influence on speech requires that it have descendants, that is, exemplars that have been produced based on its settings. Wedel notes that, due to chance elimination, most exemplars in a particular category are likely to be descended from a single exemplar, which limits the variation among the descendants. He also finds that competition for contrast maintains lexical distinctions, further mediating the effect of noise. Finally, he simulates contrast shift. This is a case in which phonetic pressure on one contrast forces it to neutralize, but competing categories can 'automatically' compensate, and broaden contrast along another dimension. Importantly, these behaviors are obtained without an explicit mechanism in place to produce them. Rather, they are complex behaviors that arise as a result of simpler, local interactions.

Several researchers have studied language change by modeling populations of artificial neural networks. Such an approach is interesting because connectionists claim that a neural network could be a full model of grammar. While other models of change often posit models of grammars and learning acquisition that do not reflect the complexity of the theories that they represent, it is fairly straightforward to use a state-of-the-art connectionist acquisition model as an agent in a simulation of language change.

Hare and Elman (1995) modeled morphological change in the history of English by modeling the acquisition of Old English morphology by a series of neural networks. The first network was trained to learn Old English verbal morphology, but training was cut off after a certain point to model imperfect learning. The output of that network was collected, and used as the input for the next network. This replicated the cumulative effect of imperfect learning over generations of speakers. In the seventh generation of networks, the verb system had simplified in much the same way that it has in present-day English. In a subsequent simulation of the maintenance of strong verbs in English, they found that their networks preserved verb forms with high type or token frequency (i.e. either those verbs from classes with many members, or simply very frequent verbs), and those in which the affixed form was predictable from phonological cues.

Livingstone (2002, 2003) reports on a series of experiments where populations of simple neural networks (perceptrons; Rosenblatt 1962) were created and allowed to interact. The networks associated a phonetic representation to a lexical representation, with a lexicon of eight words. Learning proceeded in generations, with children learning from their parents. Random noise was added to productions, to add variability to the simulation. Livingstone's primary finding was that 'dialects' – distinct phonetic representations that were localized in the social structure – emerged. This occurred even in the absence of any difference in the social standing of the individuals, contrary to the results of Nettle (1999a).

The ability of Livingstone's model to generate dialectal variation independent of social factors, contra Nettle, raises questions about which specific model parameters caused a difference in the simulation. Absent a systematic adjustment of parameters in each model, any explanation for the differences must be tentative. In Nettle's (1999a) simulations, each speaker chose to adopt one variant or another. In contrast, speakers in Livingstone's model were attempting to learn a phonetic representation, without having explicit awareness of variant pronunciations. Livingstone's model appears to accumulate noise over time (much like Hare and Elman's), while Nettle's is more the product of deliberate 'choice'. Although there is noise in Nettle's model, it is introduced in quanta: the acquisition of p rather than q, or vice versa. Nettle's model, in contrast, allows for noise that can obscure category boundaries, and subtly change pronunciations over time. A more rigorous comparison is in order, however, to determine the reason that the models behaved differently.

2 Models of Syntactic Change

One of the more influential computational models of language change was initiated by Niyogi and Berwick (1995a,b, 1996, 1997), who derived fairly simple mathematical generalizations about language change after making some simplifying assumptions. They model an infinite, perfectly mixed population. They also assume that each learner adopts a single grammar, and that the grammar is learned with a simple learning algorithm. The Triggering Learning Algorithm (TLA; Gibson and Wexler 1994) is a memoryless learning algorithm. The learner's guess about the appropriate grammar depends only on the last pertinent evidence it encountered. To offer an extreme example, the learner might encounter 100 data that indicate head-first phrase structure, and then just a single datum that indicates head-final phrase structure. It will learn head-final phrase structure, since that is the most recent evidence encountered.[6] These assumptions permit the elegant mathematical formulation of the behavior of the population. Specifically, the probability of learning a parameter setting depends on the probability of encountering a datum that is parsable only with that parametric value. Since the grammar is memoryless, the probability of acquiring a certain parameter setting is identical to the probability of encountering a datum that requires that parameter setting.

These assumptions lead to the generalization that if two grammars, G_1 and G_2, are competing, then the one that will be selected more often is the one that has the fewer sentences that can be parsed by the other grammar. Some of the sentences generated by G_1 are parsable only by G_1; others might also be parsable by G_2, however. Likewise, a proportion of the sentences generated by G_2 will also be parsable with G_1. The more sentences a grammar generates that are parsable by another grammar, the less likely a learner is to chose the source grammar. The source grammar is more likely to be chosen by a learner if the input data contain data that unambiguously identify that grammar.

As an example of this kind of reasoning, Niyogi (2002b) considers the development of English syntax from V2 to non-V2. He counts the degree-0 sentences (i.e. those without an embedded clause) that are generated by both grammars. It turns out that ambiguous sentences form a greater part of the set of non-V2 sentences, while V2 sentences are less frequently ambiguous. If it is assumed that each sentence type in a language has an equal probability of being generated, V2 grammar is always expected to prevail, contrary to the historical development of English. On the other hand, if a V2 grammar that otherwise has the same parameter settings of present-day English is compared to an otherwise identical head-final grammar, the proportion of ambiguous sentences is identical, so that no change over time is expected. These studies demonstrate how a theory of acquisition can lead to unexpected results when language change is modeled. The results are admittedly the product of simple modeling assumptions and a simple learning algorithm: one can imagine that not all sentence types have equal probabilities of being generated, or that children have more robust acquisition strategies than the TLA. Such alternatives are considered by the researchers following Niyogi and Berwick's general approach.

Yang (2000, 2003) studies acquisition-based change in a manner similar to Niyogi and Berwick, but he broadens their approach by modeling speakers that can control two grammars simultaneously. Through learning speakers determine the frequency with which each grammar will be used. To set the frequency of different variants, each learner uses a learning penalty algorithm. A grammar is penalized (i.e. used less frequently) if it encounters data that it cannot parse; this discourages the use of grammars that are ineffective. Considering the case of competing grammars G_1 and G_2, Yang defines two variables: α is the proportion of G_1 utterances that are incompatible with G_2, and β is the proportion of G_2 utterances that are incompatible with G_1. These are used to formulate the 'fundamental theorem of language change', that is, 'G_2 overtakes G_1 if $\beta>\alpha$: [that is, if] the advantage of G_2 is greater than that of G_1.' Even with the more complex acquisition mechanism, then, Yang's results are similar to those of Niyogi and Berwick.

With his expanded model, Yang analyzes the loss of V2 in Old French, a difficult case, because, as noted above, the V2 parameter is expected to be resistant to change. In Yang's analysis of Old French, the crucial observation is that the Old French was a pro-drop language. He considers the degree-0 sentences of Old French, and notes that the advantage that V2 grammars have over SVO grammars is lost when pronominal subjects are dropped. Specifically, two types of sentences that otherwise encourage the acquisition of V2 become ambiguous: OVS and XVSO. Unlike Niyogi and Berwick, Yang does not assume that all sentences of a grammar will occur with equal frequency. He rather draws on frequency counts from historical corpora to determine how frequently various sentence types might be encountered. He calculates that pro-drop diminishes the advantage of the V2 grammar to the point where a shift to SVO is inevitable. Yang also analyzes the loss of V2 in Middle English. The analysis is parallel. Middle English was V2, except that pronominal subjects immediately preceded the verb as clitics, giving rise to a verb in third position (V3). Yang follows van Kemenade (1987) in analyzing the demise of V2 as a consequence of the elimination of subject pronominal clitics, which made the (formerly) V3 constructions compatible only with an SVO analysis.

Mitchener (2005) generally follows the Niyogi and Berwick line of research, while recognizing (with Yang 2000, 2003) the need to model speakers' ability to use more than one grammar simultaneously. He represents competing grammars as a single fuzzy grammar that stochastically produces discrete grammars (similarly to probabilistic OT, as discussed above). He tests two different learning algorithms, in order to test their relative merits in explaining language change. In the 'learn-always' algorithm, a learner generates a discrete grammar, and parses an input sentence. If the chosen grammar parses the sentence correctly, then one of the parameters of the grammar (chosen at random) is rewarded so that it is favored by the fuzzy grammar in subsequent parses. The 'parameter-crucial' algorithm is similar except that, in a successful run, after a parameter is randomly selected, a hypothetical grammar is constructed that has the opposite value for that parameter, and the sentence is parsed with that grammar. If the hypothetical parse succeeds, the parameter can be judged not to have been crucial, and it is not rewarded. If the hypothetical parse fails, then the parameter is judged to be crucial, and it is rewarded. His simulations revealed interesting patterns. The 'learn-always' algorithm produced incoherent, widely varying populations. To address this, Mitchener added a bias that encouraged the fuzzy grammars toward categorical behavior rather than gradience. (The same bias added to the 'parameter-crucial' resulted in more efficient computation, and Mitchener reports that the effect this parameter had on others was minimal). Importantly, both classes of grammars acquired and lost the V2 parameter freely, a behavior that was not consistently produced in other models of this sort (but cf. the discussion of Yang's work above).[7]

Models of the type discussed above have been used to study language change for its own sake. Pearl and Weinberg (2007), however, use the model in order to test two theories of how children acquire syntax. They examine the proposals that children choose not to learn from ambiguous sentences, and that they do not learn from sentences with embedded clauses (Lightfoot 1991; Fodor 1998; inter alia). Their strategy was to implement these learning theories in a simulation of sound change, to see if the strategies bring about appropriate historical changes. They construct models with and without the filters in place, and analyze the shift from Old English OV to VO in AD 1000–1200. Along the lines of Niyogi and Berwick, they note that outcome of acquisition depends on the distribution of the data encountered. These distributions are modified when ambiguous sentences are discarded by learners. For instance, they note that because of the presence of V2 constructions in Old English, it becomes ambiguous for an SVO sentence whether the verb was base-generated before the object, or whether it was generated after the object, but subsequently raised.[8] When all of the exceptions are taken into account, the evidence favors acquisition of a VO grammar. In the simulation, this results in sigmoidal shift from OV to VO grammar, over the course of about 200 years.

A point of secondary interest in Pearl and Weinberg's simulation is the temporal interpretation of their time increment, which advances in 2-year intervals. (The point is secondary because the central result is the gross outcome of their simulation, not the fact that it matches the time course of the historical change.) A reason for the 2-year interval is not given, which leads one to suspect that it was chosen so that the time course of the simulation would match the time course of the historical change, and by extension that the time course of *any* language change could be matched by adjusting the temporal interpretation of a time step. This critique is not specific to the work of Pearl and Weinberg: Baker (2008) acknowledges that the interpretation of the time step is crucial for the interpretation of certain results of his model as well.

Nakamura et al. (2003) apply a similar model to the study of the formation of creoles. Assuming that children receive input from all language communities, they note that the crucial factor in determining the final state is the mutual intelligibility of the source languages. As a simplification, they suppose that there are three possible grammars: two source languages and a creole grammar. In the case where the source languages are entirely mutually unintelligible, but are each somewhat intelligible with a creole grammar, the rise of the creole is inevitable. This is because a speaker attempting to parse the ambient data with the grammar of one of the source languages would not be able to parse sentence from the other source language, but would be able to parse some creole sentences. Parsing the ambient data with a creole grammar allows data from all three languages to be parsed, at least in part. This gives the creole grammar an advantage. Beyond this simple state many results are possible, resulting either in the rise of a creole or not and, for cases where a creole does rise, for the creole to be dominant or non-dominant in the population. The crucial requirement for one of the source languages to become dominant is for it to be intelligible enough with the creole that children can plausibly acquire the source language from the creole data.

The work of Nakamura et al. makes several simplifying assumptions: they have a short list of potential grammars, they neglect social interactions, and they do not account for speakers' abilities to manage multiple grammars simultaneously. Nevertheless their contribution is real, in applying a model developed for other purposes to the more complex problem of creole genesis.

The research reported in this section is based on the assumption that the grammar that is most compatible with the input data is the one that will be selected. If this is the case, then the question arises as to why all languages do not have the syntax of Dyirbal, which allows free word order (Dixon 1972). If children adopt the grammar most compatible with the input, then it seems inevitable that they should eventually converge on the grammar that poses fewest restrictions on the input. To pose the problems in the formalism of Yang (2000, 2003), if one of the grammars (say G_1) allowed free word order, then β would equal 0, and G_1 would always be learned. Assuming that speech errors are confined in some way, this might not be expected to occur in a single generation, but it should be observed over time.

This specific problem is an instance of a more general problem noted by Briscoe (2000). If innate biases are built into an acquisition model, then we might expect to see languages optimizing over time to conform to the acquisition model. This predicts convergence of languages, rather than the observed divergence of languages. Following the above example, if the acquisition bias is to choose the most parsable grammar, we would expect to see all grammars evolve to be inclusive of more word orders. Briscoe (2000) proposes that this problem might be dealt with by acknowledging multiple competing forces in language acquisition. That is, instead of identifying a single factor as the driving force selecting grammars (such as ability to parse many sentences), several factors might be at play, which together could define many locally 'ideal' parameter settings. Conceptually, the proposal seems to be on the right track, but the difficulty lies in identifying precisely what the competing motivations are, and what happens when they are in conflict.

3 Future Directions

Studies of language change have proved fruitful in uncovering hidden consequences of theories and facts. For instance, simulations by Niyogi and Berwick uncovered previously unknown (often incorrect) predictions of the Triggered Learning Algorithm. Nettle (1999a) discovered the significance that social factors play in change, at least for cases where two variants are in competition. These findings are valuable, because they indicate areas that require more careful study, as indeed they already have. For instance, the problems noted by Niyogi and Berwick (1996) in the loss of V2 were addressed by the work of Yang (2000, 2003).

A number of researchers have noted that the elegant mathematical formulations of language change offered by Niyogi and Berwick (1995a, et seq.) are lost when the model's basic assumptions, such as infinite population size or non-stochastic learning, are changed (Clark 1996; Briscoe 2000; Niyogi 2002b; Livingstone 2003). If a mathematical model of language change is unavailable, researchers must use Monte Carlo methods, performing large number of simulations in order to determine the effects of different model parameters. Each of the studies presented in this article involves many model parameters, and each reported selectively on the effect of varying these parameters. A common practice in reporting results is to present graph of a typical run, or to comment on the qualitative result of varying a certain parameter. While this approach can be used to give a concise and accurate description of runs, it is not a rigorous way to report the results of simulations. Ideally, a report on a model would include the effect of each model parameter on the outcome of the simulation.[9] The most straightforward way to do this would be a regression analysis, with each model parameter as an independent variable, and the pertinent outcome variable the dependent variable. Increased availability of computing power makes sampling the entire parameter space a practical task. Adding statistical rigor to computational studies might go a long way in moving the discussion from dissertations, book chapters, and conference proceedings – where it has almost exclusively take place – into peer-reviewed journals.

One difficulty in evaluating models is that they are frequently incommensurate. The models of Nettle (1999a) and Livingstone (2003) differed in whether they could produce change without social variation, but the models differed in so many respects that it is not clear which difference (or set of differences) was crucial. Was it the presence of multiple words in Livingstone (2003)? The use of a neural net as opposed to Nettle's (1999a) equation from Social Impact Theory? Closer integ- ration between models, and more systematic studies of how variation in parameters produces different results, will be a substantial benefit to the field.

Several authors have noted the importance of network structure in simulations of language change (Lee et al. 2005; Ke et al. 2007; Magué 2007). For instance, Ke (2007) models individuals in a fully connected network, using similar model parameters to Nettle (1999a). She finds that if each interaction with another individual who has adopted a change carries a certain probability of adopting the change, change propagates through the community in a sigmoidal fashion. With a more sparsely connected network, change proceeded as before, but at a lower rate. Different network architectures also made a difference in the outcome of simulations (which are more fully discussed in Ke's article). Differences introduced by different network structures have not been studied systematically for most models, but clearly the structure of the network has a bearing on the outcome of the simulation. Future work should take this into account by either matching the structure of a real-life network (if data are available), or by running simulations with multiple network structures to assess the role that the structure of the social network has in determining the outcome of a simulation.

4 Conclusion

It is widely thought that language change is driven by imperfect acquisition. Niyogi (2002a) claims that, 'perfect acquisition would imply perfect transmission.' Studies of language change indicate that, although the situation is far more complex than this, since individual speakers do make choices about their linguistic habits beyond childhood, language acquisition is nevertheless important to the study of change (Labov 1994, 2001). Consequently, simulations of language change will benefit as models of acquisition improve and provide better agents to incorporate into models. Conversely, models of acquisition will benefit as they take into account not only child developmental data, but also the consequences of their acquisition models for the historical development of a language; Pearl and Weinberg's (2007) work is an innovative beginning to these types of studies. The beginnings of closer integration are also evident in the neural network literature (e.g. Hare and Elman 1995). But the possibilities have yet to be explored in other modeling efforts. A replication of the Hare and Elman (1995) study with a probabilistic OT grammar (e.g. Boersma 1997; Boersma and Hayes 2001) may be a reasonable next step.

Zuidema (2003) notes that models of syntactic parameter setting never involve a full grammar, but only the two or three parameters that are most easily interpretable by linguists. This criticism could be made of many models of language change as well. Studies of sound change could easily incorporate larger lexicons, modeling more changes simultaneously, than currently is done. There is no reason not to expand these efforts, to test the quality of the current ideas against more challenging data.

Understanding the role that acquisition errors play in language change is a significant problem. Acquisition errors are the crucial seed for language change (Nettle 1999a; Livingstone 2003), providing variability from which subsequent changes are drawn. Models of change depend on models of acquisition errors, then. As acquisition models become more complex, and better able to represent children's performance, they will offer theories of change more specific kinds of errors, which will aid understanding of the specific role that acquisition errors play in language change.

Computational models of language change, which have a relatively recent advent, are quickly becoming standard tools in theorization about the origins and causes of change. There is every reason to be optimistic about the subsequent development of these fields. As computers become faster and as more corpora become available, researchers will have ever-increasing freedom to develop and test models of change.

Short Biography

Adam Baker is a PhD candidate in the Department of Linguistics at the University of Arizona. His primary theoretical interests are the locus of explanation in for sound patterns (particularly with regard to the split between synchronic and diachronic explanation), the phonetic and social motivations for sound change, and the representation of speech sounds in the brain. In pursuing these interests, he has conducted phonetic experiments with speakers of English, French, and Turkish. His dissertation is titled 'A biomechanical model of the human tongue for understanding speech production and other lingual behaviors'. He holds a BA and MA in linguistics from the University of Arizona.

Notes

* Correspondence address: Adam Baker, Department of Linguistics, University of Arizona, Douglass Building, Room 200E, Tucson, AZ 85721, USA. E-mail: tabaker@u.arizona.edu.

[1] The article does not review the vast literature on the application of computational methods to the study of the origins of the language faculty itself. This is commonly called 'language evolution', although using the term is often imprecise, since 'evolution' informally refers to any change through time. Many authors note potential analogs between the evolution of the language faculty and the historical changes that languages undergo (e.g. de Boer 2000; Nowak and Komarova 2001), and some study both phenomena with similar techniques (e.g. Briscoe 2002; Livingstone 2003; Wang et al. 2004; Wang and Minett 2005). Nevertheless, the fields are fundamentally different: models of language change begin with linguistic grammars, while evolutionary models do not (see Cangelosi and Parisi 2002 for further discussion). See Nowak et al. 2002, Christiansen and Kirby 2003, or Brighton et al. 2005 for recent reviews of the language evolution literature. To these studies can be added two other classes of computational studies, which model broad cross-linguistic tendencies without modeling specific languages (e.g. Kirby (1996, 1997); Jäger (2003) and the earlier applications of the de Boer (1999) model), and

the relatively new work in the modeling the competition of languages for speakers to understand language shift phenomena (Abrams and Strogatz 2003; Stauffer and Schulze 2005; de Oliveira et al. 2006; Schulze and Stauffer 2006; Stauffer et al. 2006).

[2] Hence, the common criticism noted by Nettle (1999a: 103) that modelers simply insert whatever behavior they desire into a model, run the simulation, and claim to have modeled the phenomenon.

[3] Curiously, Nettle considered this to be a negative feature of the model. To the contrary, it might rather be interpreted as the simulation of a split in the speech community, which is an interesting result.

[4] Changes in language often do not proceed linearly, but are instead sigmoidal (Bailey 1973; Kroch 1989; Labov 1994, 2001). Many modeling papers make sigmoidal propagation of a change through a population a specific deseridatum for the model.

[5] de Boer and Kuhl (2003) subsequently showed that learning positions of adult vowel categories from infant-directed speech was more successful than learning those categories from adult-directed speech.

[6] Niyogi and Berwick's model is described in several papers; Niyogi (2004) provides the most accessible introduction.

[7] For a mathematically rigorous approach, the reader may be referred to Mitchener (2003), a dissertation in mathematics devoted to the investigation of language change from a game-theoretic perspective.

[8] Note that this is structural ambiguity, not ambiguity as to whether the string can be parsed by one grammar or another.

[9] Of the models surveyed here, Nettle (1999a) comes closest to this ideal; Choudhury et al. (2006) might be taken as an example of an extremely complex model with no report of the result of varying the model parameters.

Works Cited

Abrams, Daniel, and Steven H. Strogatz. 2003. Modelling the dynamics of language death. Nature 424.900.

Bailey, Charles-James N. 1973. Variation and linguistic theory. Washington, DC: Center for Applied Linguistics.

Baker, Adam. 2008. Addressing the actuation problem with quantitative models of sound change. Proceedings of the 31st Annual PLC. Penn Working Papers in Linguistics 14.1.

de Boer, Bart. 1999. Self-organisation in vowel systems. PhD thesis, Vrije Universiteit Brussel.

——. 2000. Self organization in vowel systems. Journal of Phonetics 28.441–65.

——. 2002. Evolving sound systems. Simulating the evolution of language, ed. by Angelo Cangelosi and Domenico Parisi, 3–28. New York, NY: Springer-Verlag New York, Inc.

——. 2003. Conditions for stable vowel systems in a population. Advances in Artificial life: Proceedings of the 7th European Conference on Artificial Life (ECAL 2003), Dortmund, Germany. Lecture Notes in Computer Science 2801, ed. by Wolfgang Banzhaf, Thomas Christaller, Peter Dittrich, Jan T. Kim and Jens Ziegler, 415–24. Berlin, Germany: Springer.

de Boer, Bart, and Patricia K. Kuhl. 2003. Investigating the role of infant-directed speech with a computer model. Acoustics Research Letters Online 4.129–34.

Boersma, P., and B. Hayes. 2001. Empirical tests of the gradual learning algorithm. Linguistic Inquiry 32.45–86.

Boersma, Paul. 1997 How we learn variation, optionality, and probability. Institute of Phonetic Sciences, University of Amsterdam Proceedings 21, 43–58.

Brighton, Henry, Kenny Smith, and Simon Kirby. 2005. Language as an evolutionary system. Physics of Life Reviews 2.177–226.

Briscoe, Edward J. 2000. Evolutionary perspectives on diachronic syntax. Diachronic syntax: models and mechanisms, ed. by S. Pintzuk, G. Tsoulas and A. Warner, 75–108. Oxford, UK: Oxford University Press.

——. 2002. Introduction. Linguistic evolution through language acquisition, ed. by Edward J. Briscoe, 1–21. Cambridge, UK: Cambridge University Press.

Cangelosi, Angelo, and Domenico Parisi. 2002. Computer simulation: a new scientific approach to the study of language evolution. Simulating the evolution of language, ed. by Angelo Cangelosi and Domenico Parisi, 3–28. New York, NY: Springer-Verlag.

Choudhury, Monojit, Anupam Basu, and Sudeshna Sarkar. 2006. Multi-agent simulation of emergence of Schwa deletion pattern in Hindi. Journal of Artificial Societies and Social Simulation 9. <http://jasss.soc.surrey.ac.uk/9/2/2.html>.

Christiansen, Morten H., and Simon Kirby. 2003. Language evolution: consensus and controversies. Trends in Cognitive Sciences 7.300–7.

Clark, R. A. J. 1996. Internal and external factors affecting language change: a computational model. PhD thesis, MSc Dissertation, University of Edinburgh.

Culicover, Peter, Andrzej Nowak, and Wojciech Borkowski. 2003. Linguistic theory, explanation and the dynamics of grammar. The nature of explanation in linguistic theory, ed. by J. Moore and M. Kolinsky, 83–104. Stanford, CA: CSLI Publications.

Dixon, R. M. W. 1972. The Dyirbal language of North Queensland. London: Cambridge University Press.

Dras, M., and K. D. Harrison. 2002. Emergent behavior in phonological pattern change. Proceedings of Artificial Life VIII.390–393.

Fodor, Jerry D. 1998. Unambiguous triggers. Linguistic Inquiry 29.1–36.

Gibson, E., and K. Wexler. 1994. Triggers. Linguistic Inquiry 25.407–454.

Hare, Mary, and Jeffrey L. Elman. 1995. Learning and morphological change. Cognition 56.61–98.

Harrison, K. D., M. Dras, and B. Kapicioglu. 2002. Agent-based modeling of the evolution of vowel harmony. Proceedings of the Northeast Linguistic Society 32, ed. by Mako Hirotani. Amherst, MA: GLSA.

Jäger, Gerhard. 2003. Simulating language change with functional OT. Language Evolution and Computation, Proceedings of the Workshop at the 15th European Summer School on Logic Language and Information, ed. by Simon Kirby, 52–61. <http://www3.isrl.uiuc.edu/~junwang4/langev/localcopy/pdf/jager03ESSLLI.pdf>.

Ke, Jinyun. 2007. Complex networks and human language. <http://www.isrl.uiuc.edu/amag/langev/paper/ke07complexNetworkLanguage.html>.

Ke, Jinyun, Tao Gong, and William S-Y Wang. 2007. Stochastic optimality theory, local search, and bayesian learning of hierarchical models. <http://dingo.sbs.arizona.edu/~yinglin/Lin_hierarchical.pdf>.

van Kemenade, A. 1987. Syntactic case and morphological case in the history of English. Dordrecht, The Netherlands: Foris.

Kirby, Simon. 1996. Function, selection and innateness: the emergence of language universals. PhD thesis, University of Edinburgh.

——. 1997. Competing motivations and emergence: explaining implicational hierarchies. Language Typology 1.5–32.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. Language Variation and Change 1.199–244.

Kuhl, Patricia K., Jean E. Andruski, Inna A. Chistovich, Ludmilla A. Chistovich, Elena V. Kozhevnikova, Viktoria L. Ryskina, Elvira I. Stolyarova, Ulla Sundberg, and Francisco Lacerda. 1997. Cross-language analysis of phonetic units in language addressed to infants. Science 277.684–686.

Labov, William. 1994. Principles of linguistic change, volume 1: Internal factors. Oxford, UK: Blackwell.

——. 2001. Principles of linguistic change, volume 2: social factors. Oxford, UK: Blackwell.

Latané, B. 1981. The psychology of social impact. American Psychologist 36.343–56.

Lee, Y., T. C. Collier, E. P. Stabler, and C. E. Taylor. 2005. The role of population structure in language evolution. Proceedings of the 10th International Symposium on Artificial Life and Robotics <http://www.linguistics.ucla.edu/people/stabler/LeeEtAl2005theRole.pdf>.

Lightfoot, David. 1991. How to set parameters. Cambridge, MA: MIT Press.

Livingstone, Daniel. 2002. The evolution of dialect diversity. Simulating the evolution of language, ed. by Angelo Cangelosi, and Domenico Parisi, 99–118. New York, NY: Springer-Verlag.

Livingstone, Daniel J. 2003. Computer models of the evolution of language and languages. PhD thesis, University of Paisley.
Magué, Jean-Philippe. 2007. On the importance of population structure in computational models of language change. <http://www.ling.upenn.edu/Events/PLC/plc31/abstracts/11636072276501.pdf>.
Mitchener, W. Garrett. 2003. A mathematical model of human languages: the interaction of game dynamics and learning processes. PhD thesis, Princeton University.
——. 2005. Simulating language change in the presence of non-idealized syntax. Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition, 10–9. Ann Arbor, MI: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W05/W05-0502>.
Nakamura, Makoto, Takashi Hashimoto, and Satoshi Tojo. 2003. Creole viewed from population dynamics. Language Evolution and Computation Workshop/Course at the 15th European Summer School on Logic Language and Information, Vienna, 25–29 August, ed. by S. Kirby, 95–104. <http://www3.isrl.uiuc.edu/~junwang4/langev/localcopy/pdf/nakamura03esslli.pdf>.
Nettle, Daniel. 1999a. Using social impact theory to simulate language change. Lingua 108.95–117.
——. 1999b. is the rate of linguistic change constant? Lingua 108.119–36.
Nichols, J. 1990. Linguistic diversity and the first settlement of the new world. Language 66.475–521.
——. 1992. Linguistic diversity in space and time. Chicago, IL: University of Chicago Press.
Nigg, Benno M. 1994. General comments about modelling. Biomechanics of the musculo-skeletal system, ed. by B. M. Nigg and W. Herzog, 367–79. Sussex, UK: John Wiley & Sons.
Niyogi, P., and R. C. Berwick. 1995a. A dynamical systems model for language change. Tech Rep AI Memo-1515, MIT.
——. 1995b. The logical problem of language change. Technical report, AI Memo-1516. AI Lab, MIT.
——. 1996. A language learning model for finite parameter spaces. Cognition 61.161–93.
——. 1997. Evolutionary consequences of language learning. Linguistics and Philosophy 20.697–719.
Niyogi, Partha. 2002a. The computational study of diachronic linguistics. Syntactic effects of morphological change, ed. by David Lightfoot, 351–65. Cambridge, UK: Cambridge University Press.
——. 2002b.Theories of cultural evolution and their application to language change. Linguistic evolution through language acquisition, ed. by Edward J. Briscoe, 205–33. Cambridge, UK: Cambridge University Press.
——. 2004. Phase transitions in language evolution. Variation and universals in biolinguistics, ed. by L. Jenkins. Oxford, UK: Elsevier.
Nowak, M. A., and N. L. Komarova. 2001. Towards an evolutionary theory of language. Trends in Cognitive Sciences 5.288–95.
Nowak, Martin A., Natalia L. Komarova, and Partha Niyogi. 2002. Computational and evolutionary aspects of language. Nature 417.611–7.
Ohala, John J. 1993. Coarticulation and phonology. Language and Speech 36.155–70.
de Oliveira, Viviane M., Paulo R. A. Campos, M. A. F. Gomes, and I. R. Tsang. 2006. Bounded fitness landscapes and the evolution of the linguistic diversity. Physica A 368.257–61.
Pearl, Lisa, and Amy Weinberg. 2007. Input filtering in syntactic acquisition: answers from language change modeling. Language Learning and Development 3.43–72.
Pierrehumbert, Janet B. 2002. Exemplar dynamics: Word frequency, lenition and contrast. Frequency and the emergence of linguistic structure, ed. by Joan Bybee and Paul Hopper, 323–418. Amsterdam, The Netherlands: John Benjamins.
Rosenblatt, Frank. 1962. Principles of neurodynamics. New York, NY: Spartan Books.
Schulze, Christian, and Dietrich Stauffer. 2006. Recent developments in computer simulations of language competition. Computing in Science & Engineering 8.60–7.
Shen, Zhongwei. 1997. Exploring the dynamic aspect of sound change. Journal of Chinese Linguistics Monograph no. 11.

Stauffer, D., C. Schulze, F. W. S. Lima, S. Wichmann, and S. Solomon. 2006. Non-equilibrium and irreversible simulation of competition among languages. Physica A 371.719–24.

Stauffer, Dietrich, and Christian Schulze. 2005. Microscopic and macroscopic simulation of competition between languages. Physics of Life Reviews 2.89–116.

Wang, William S-Y., J. Y. Ke, and James W. Minett. 2004. Computational studies of language evolution. Computational linguistics and beyond, ed. by W. Lenders and C. R. Huang, 65–106. Nankang, Taipei: Institute of Linguistics, Academia Sinica.

Wang, William S-Y., and James W. Minett. 2005. The invasion of language: emergence, change and death. Trends in Ecology and Evolution 20.263–69.

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. The Linguistic Review 23.247–74.

Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. Directions for historical linguistics, ed. by W. Lehmann and Y. Malkiel, 95–195. Austin, TX: University of Texas Press.

Yang, Charles. 2003. Knowledge and learning in natural language. Oxford, UK: Oxford University Press.

Yang, Charles D. 2000. Internal and external forces in language change. Language Variation and Change 12.231–50.

Zuidema, Willem. 2003. Modeling language acquisition, change and variation. Language Evolution and Computation Workshop at the 15th European Summer School on Logic Language and Information, Vienna, 25–29 August, ed. by Simon Kirby, 32–41. <http://www3.isrl.uiuc.edu/~junwang4/langev/localcopy/pdf/zuidema03acquisitionChangeVariation.pdf>.