Propagation of change: computational models

Henri Kauhanen

N.B. This is a **preprint** of an article accepted for publication in the *Wiley-Blackwell Companion to Diachronic Linguistics*. It does not reflect changes made during the peer review process. When citing, please refer to the published version.

Abstract

Linguistic propagation occurs when one linguistic variant replaces another, or several others, over time. Evolutionary dynamics provides a formal framework in which to describe and model processes of propagation; employing this framework, we provide an overview of computational models of propagation which have been proposed in the literature. These range from simple phenomenological models to more complex models that take geographical space or social structure into account; both constant-fitness models and models involving frequency-dependent selection are discussed. Special emphasis is paid on the epistemological role mechanistic models play as formal apparatuses which can derive (not just describe) large-scale regularities of language variation and change from underlying first principles.

Keywords: computational models, evolutionary dynamics, language change, mathematical models, propagation

1 Introduction: computational models and the evolutionary dynamics of language

Linguistic mutations—innovations and "mistakes" which have the potential to propagate through the speech community and give rise to community-level change—occur all the time: in first language acquisition, in second language acquisition, and in the linguistic production of speakers whose grammatical systems have already stabilized. Only a minority of these innovations, however, actually manage to propagate—at any given point in time, *most* aspects of a language are *not* undergoing change. What are the factors governing these dynamics? When innovations *do* succeed, *why* do they succeed? What role is played in this by random events, such as the stochastic nature of speaker-to-speaker interactions? Conversely, to what extent is the propagation of linguistic mutations governed by deterministic factors and hence amenable to deductive–nomological explanation and prediction?

One effective way of approaching these questions is through the use of mathematical modelling techniques. Formally specified models have a number of well-known benefits: they invite the researcher to spell out all background assumptions and hypotheses explicitly, they lead to at least qualitative and often quantitative predictions, and they can often be applied, as a sort of *in silico* laboratory, in situations in which other kinds of approaches are ineffective or outright impossible. More and more models of this kind are becoming available in the vast and growing body of literature on language variation and change, and numerous intriguing conceptual connections with neighbouring fields, in which dynamical processes play equally central roles, await to be fully explored. It is the purpose of this chapter to give the reader a taste of this breadth of modelling approaches, while at the same time offering a programmatic account of a cultural-evolutionary approach to variation and change.

The phrases 'formal model', 'mathematical model' and 'computational model' are all used in the literature, sometimes interchangeably, sometimes not. Here, I use the term 'computational model' in a wide sense, to refer to any formally specified model, whether it is studied using analytical mathematical methods or computer simulations. Examples of different types of computational models, specified on different levels of complexity and studied using different techniques, will be given throughout the chapter.

Although it is important to highlight the differences between different models, it is equally important to stress their commonalities. Formal models of language variation and change are unified by the fact that they attempt to describe and explain the *evolutionary dynamics* of linguistic systems. This implies that such models view language change as a (cultural) evolutionary process in which 'different ways of saying the same thing' compete in populations of language users. These processes may vary widely from one situation to the next: for instance, they may be selectively neutral in some cases but involve differential fitness in others. They share the following central aspects, however: (i) linguistic changes occur through the replication of linguistic variants in language communities, and (ii) those changes, when viewed on a macroscopic level, are *emergent* products of numerous interactions that occur on the lower, fundamental level of language acquisition and linguistic interaction. These assumptions are important, because they have epistemological consequences. For instance, on this view, various directionality properties of change processes (such as have been extensively documented e.g. in the study of grammaticalization; Hopper and Traugott 1993; also see Unique ID WBCDL026) cannot be taken as epistemologically primary; rather, they should properly be viewed as *explanantia*, something to be explained. It is the language dynamicist's task to show how such macroscopic, population-level and possibly very long-term developments emerge from an underlying, ontologically primary level of individual cognition and social interaction.

2 Propagation of change: three problems

It is useful to view the propagation of linguistic innovations in populations of language users through the lens of three problems which are intimately related.

The first problem concerns the emergence of innovations or "linguistic mutations" in the first place. Classically, this has been discussed in the context of the *actuation problem* as formulated by Weinreich, Labov, and Herzog (1968): why do changes occur in some languages at some times, but not in others or at other times? Despite nearly five decades of work, it has been asserted that the "actuation riddle" remains essentially unsolved (Walkden 2017, 420). According to some, linguistic change cannot so much be explained as observed (Lass 1980, 1997)—a rather pessimistic conclusion for historical linguistics as a whole. Other researchers have pointed to possible solutions to the actuation problem, usually stressing the importance of bifurcations—large qualitative changes to the behaviour of a complex system in response to minute variation in some of its control parameters—in explaining how "sudden" shifts sometimes (but not always) occur (Niyogi 2006; Kirby and Sonderegger 2013; Sóskuthy 2015; also see Unique ID WBCDL001 and Unique ID WBCDL063).

Contrasting linguistic with biological evolution, Nettle (1999) identified a related challenge, which he dubbed the *threshold problem*. In biological evolution, the transmission of genetic information involves one or two parents only; consequently mutations, once they have occured, are likely to be inherited. In the transmission of linguistic information, by contrast, each language learner has a larger number of "cultural parents", some of whom, in fact, come from the learner's own peer group. Innovative forms are by definition in the minority initially, and they thus have a harder time finding their way into the learner's eventual linguistic repertoire. This has two consequences for modelling work. Firstly, any formal account of the propagation of a linguistic innovation must explain how the innovation manages to cross the threshold of propagation. Less obviously, but equally importantly, we should be skeptical of models which predict no such threshold to exist; on which more below.

In order to propagate, a linguistic mutation must overcome the actuation and threshold problems. These are necessary, but not sufficient, conditions: in general, we must also ask what keeps a change going, once it is underway. Most scholars would agree that, in some intuitive sense, a change is not really a change unless it goes to completion or near-completion: failed changes are changes in name only. This could be termed the *sustain problem*: why are changes sustained (carried through) in some languages at some times, but not in others or at other times? What makes innovative linguistic variants successful?

The actuation, threshold and sustain problems form the hard core of the explanation of linguistic change: answering the questions posed by them answers why a linguistic mutation appeared, why it rose above the threshold of propagation, and why its propagation was successful in the end. In the rest of this chapter, I will explore how models of change formulated in mathematical terms and studied through computational means can help us make sense of the three problems.

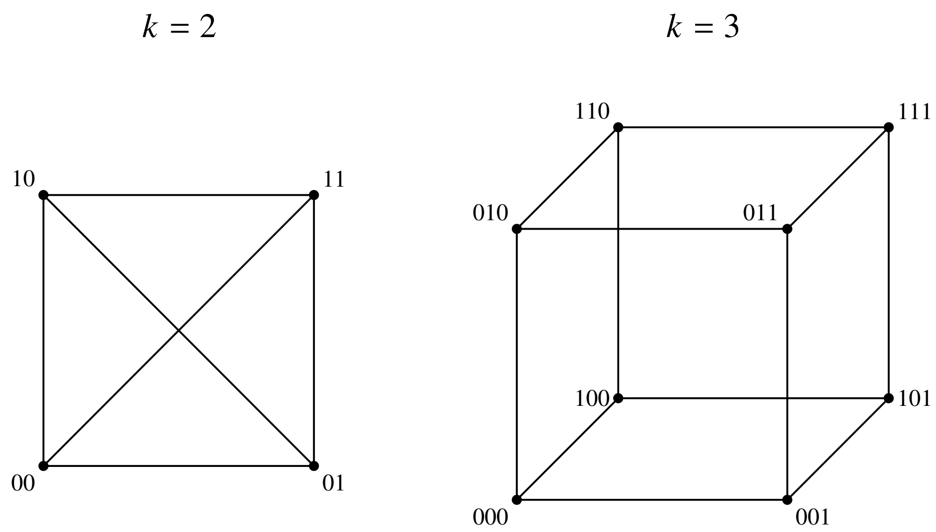

Figure 1. The configuration spaces for two (left) and three (right) binary variables. Each point is a possible language. Pairs of languages are connected by lines showing possible language changes (for clarity, changes involving flipping the value of more than one variable simultaneously are not shown in the $k = 3$ case).

3 The configuration and variation spaces of language: what counts as change?

To talk coherently about language change, we need precise definitions of both *what* changes and what is the space of *possible* change processes in language. Language change is an example of cultural evolution: a language changes when the distribution of linguistic variants across a speech community changes through the differential replication of some variants at the expense of others. To make this precise, we may imagine a hypercube whose various sides represent all the possible linguistic variables that human cognition makes available. Each point in this hypercube defines a possible language; each actually attested language is likewise a point, or perhaps a cloud of points, in this space, and either stable or moving about in the space.

To illustrate, consider a toy world in which human cognition allows k linguistic variables, all of them binary. The number of possible languages is then 2^k, and we can arrange them into a k-dimensional structure in which each of the possible languages is represented as a string $a_1 \dots a_k$ where $a_i \in \{0, 1\}$ supplies the value of the ith variable. Figure 1 illustrates this for $k = 2$ and $k = 3$. In particular, change in one or more variables now becomes equated with traversing a link between two points in this space. For instance, in the $k = 3$ case, changing the value of any one variable (from 0 to 1 or vice versa) moves us along one of the edges of the cube; changing the value of two variables at a time moves us along one of the faces of the cube; and changing the values of all three variables at once takes us through the interior of the cube.

This defines the set of possible languages and the set of possible transitions between different types of language; I will call it the *configuration space* of human language in what follows (cf. the notion of *sequence space* in biological contexts; Nowak 2006). Note that not every point of the space need ever be occupied in reality, nor need all logically possible transitions be empirically attested (cf. Unique ID WBCDL006, Unique ID WBCDL026). In fact, linguistic theory may very well place strong constraints on the set of possible languages as well as the set of possible transitions for instance in terms of variable hierarchies (parameter hierarchies in generative terms) or directionality constraints.

The configuration space affords us a geometric view of language, but it isn't enough to describe variation and change. In a change event, a language "hops" from one point in the configuration space to another; the configuration space has nothing more to say about how this happens. But the question of the propagation of change, properly construed, is precisely the question of *how* a language moves from one point in the configuration space to another.

Since languages are spoken (or signed) by individuals, and since in a language community different numbers of individuals may speak different varieties or may speak them with different probabilities, we need some notion of the "amount" of a variant that obtains in a population of speakers. That amount will be called the *abundance* of the variant; mathematically, it is a real number that must satisfy two constraints.

> **Definition (abundance).** Let $\mathbb{C}$ be a configuration space which makes available n (e.g. $n = 2^k$) variants $V_1, \ldots, V_n$. Then an *abundance vector* for $\mathbb{C}$ is any vector $\mathbf{x} = (x_1, \ldots, x_n)$ of real numbers which satisfies the following two conditions:
>
> 1. $x_i \geq 0$ for all i,
> 2. $x_1 + \cdots + x_n = 1$.
>
> The number x_i is called the *abundance* of variant V_i.

Familiar notions of the "amount" of a variant such as the proportion of speakers speaking the variant, the probability of a speaker employing a variant, or the relative frequency of that variant in a corpus of text, satisfy this definition; however, the definition is general enough to allow other possibilities as well. It should also be noted that abundance is a *local* quantity: the abundance of a variant in one population of speakers may differ from that in another population.

Every configuration space has a related space which I will call its associated *variation space*; this variation space consists of all possible abundance vectors for the configuration space. If the configuration space allows n possible variants, then the dimension of its associated variation space is also n. However, since the abundances necessarily sum to unity, that is to say, since $x_1 + \cdots + x_n = 1$, the variation space effectively has one fewer degree of freedom and can be represented as an $(n-1)$-dimensional projection. (Mathematically, for a configuration space of n variants, the corresponding variation space is the $(n-1)$-simplex.) Figure 2 illustrates.

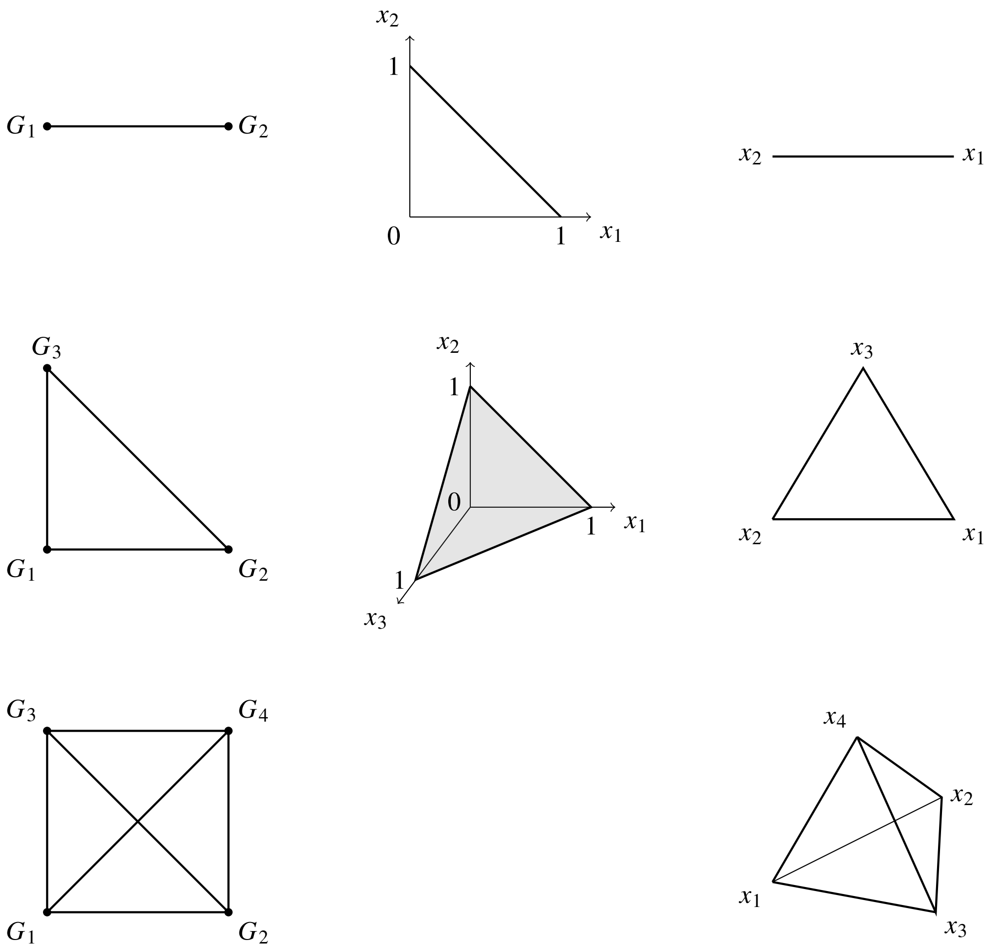

Figure 2. From configuration to variation space. The leftmost column shows configuration spaces for $n = 2, 3, 4$ languages. For each configuration space, the corresponding variation space is a subset of the n-dimensional Euclidean space subject to the requirements given in the definition of abundance (middle column). For $n = 2$, this is a straight line; for $n = 3$, it is a plane; for $n = 4$, the variation space is a three-dimensional object whose embedding four-dimensional space cannot be easily visualized. Since an n-dimensional variation space only has $n - 1$ degrees of freedom, the variation space is in fact an $(n - 1)$-dimensional manifold, a so-called simplex (rightmost column).

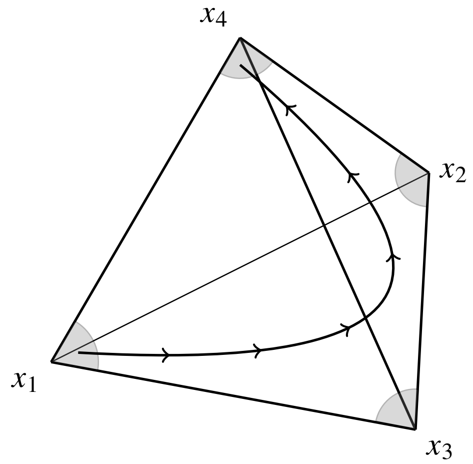

Figure 3. For each variant V_i, the delta-dominance set $D_i(\delta)$ (for some chosen dominance level $\delta > 0$) forms a local neighbourhood of the pure state in which V_i dominates fully, illustrated here as the grey regions. Propagation is said to occur when the abundance vector—a point in the variation space—travels from one delta-dominance set to another, illustrated here as the curved trajectory from $D_1(\delta)$ to $D_4(\delta)$.

Each point in the variation space (each abundance vector) represents one possible *state* of the speech community. States $\mathbf{x}$ which satisfy $x_i = 1$ for some i will be called *pure states*. All other states are *states of variation*. Since idiolectal variation combined with a highly dominant variant is often the norm in real language communities, it is useful to have a concept that refers to states which are situated close to one of the pure states yet allow a small amount of variation. I shall call a state *delta-dominant* at the level δ with respect to the ith variant if $x_i > 1 - \delta$, where $\delta > 0$ is a small number, e.g. $\delta = 0.01$ (cf. Kauhanen 2017). The set of delta-dominant states of the ith variant will be symbolized $D_i(\delta)$ (topologically, this is the open δ-ball around the vertex $\mathbf{x}$ which satisfies $x_i = 1$, intersected with the variation space).

Let $\mathbf{x}(t)$ denote the state of the speech community at time t. We can now state precisely what we mean by the propagation of an innovative variant i (see also Figure 3).

> **Definition (propagation).** Let $\mathbf{x} \in \mathbb{V}$ where $\mathbb{V}$ is a variation space. Then variant i *propagates* in $\mathbb{V}$ if and only if a variant $j \neq i$ exists such that $\mathbf{x}(t) \in D_j(\delta)$ but $\mathbf{x}(t') \in D_i(\delta)$ for some times t, t' with $t' > t$, for some suitable dominance level $\delta > 0$.

What constitutes a suitable dominance level must depend on the specifics of the application at hand; we will see examples later.

The total configuration and variation spaces of human language are high-dimensional, in fact, astronomical. Roberts and Holmberg (2010, 30) suggest it is "very likely that the number of parameters [i.e. linguistic variables] is in the

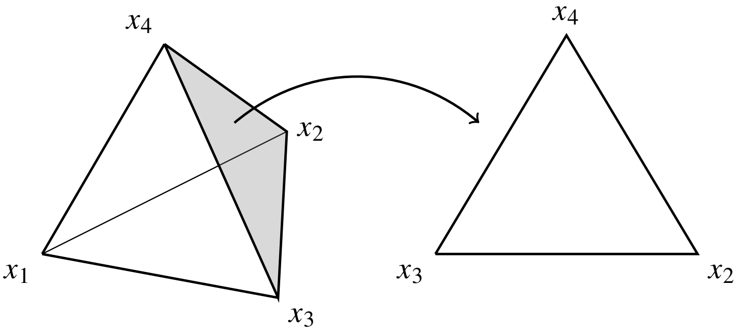

Figure 4. The Dissection Principle illustrated. Choosing any subset of pure states from a variation space (here, the pure states corresponding to the second, third and fourth variant), the subspace spanned by those states forms another variation space of lower dimensionality.

hundreds, and at least possible that it is in the thousands". Assuming every parameter to be binary for simplicity, this implies between 2^{100} and 2^{10000} distinct possible human languages, if all parameter value combinations are thought to be possible. Even if some logically possible languages are not grammatically possible, the number is still astronomical: note that even the modest lower-bound estimate of 2^{100} equals roughly 10^{30}, i.e. 1 followed by 30 zeroes. This implies that the total configuration and variation spaces are of no practical use in modelling language variation and change. More generally, if the evolution of any given linguistic variable may depend on the evolution of any other linguistic variable, so that the dimensionality of the problem may not be reduced, then *any* scientific study of language variation and change is rendered impossible. It is thus a practical necessity to "carve out" areas of the variation space for study which would appear to be governed chiefly by their own dynamics, without interference from other areas. For instance, we may suppose syntax to be largely—though perhaps not entirely—independent of phonetics, and hence study the two in separation outside of particular case studies in which such a connection or "interference" is plausible (such as when a process of phonetic reduction causes morphological erosion, ultimately leading to a restructuring of syntax). This motivates the following

> **Dissection Principle.** In focusing on the dynamics of a specific linguistic phenomenon, we may "dissect" the relevant part of the variation space, leaving the rest outside the analysis. If $X \subseteq \mathbb{V}$, then a new (dissected) variation space X is formed by the normalization $\mathbf{x}/\sum_{\mathbf{y}\in X} \mathbf{y}$ for all $\mathbf{x} \in X$.

Geometrically, each subsimplex of the original variation space is also a simplex and thus a variation space; see Figure 4.

4 The shape of propagation: a phenomenological equation

One way of responding to the sustain problem is to maintain that, much like different genomes are differently adapted to any given biological environment, different linguistic variants may also have functional differences in a given linguistic environment. The successful propagation of a variant is then explained by its greater functional advantage, when compared to its competitors. This way, predictions become available: if the advantage of one variant is greater than that of another, then, all else being equal, we should expect the abundance of the former to increase and the abundance of the latter to decrease over time.

Suppose we have reason to believe that, in some historically relevant situation, the dynamics of two variants are sufficiently isolated from correlations with further variants, so that the Dissection Principle may be invoked to limit consideration to the study of the dynamics of a single binary variable. Let $x = x_1$ refer to the abundance of the first variant; the abundance of the second variant is then $1 - x$ ($= x_2$). Suppose that each variant carries with it a *fitness*, which for the time being is simply a real number. Let f_1 stand for the fitness of the first and f_2 to the fitness of the second variant. Moreover, let

(1) $$\varphi(x) = x_1 f_1 + x_2 f_2 = x f_1 + (1 - x) f_2$$

denote the *average fitness* that obtains in the speech community, i.e. the sum of the two fitness quantities weighted by the abundances of the respective variants.

To reproduce the intuition that the fitter variant ought to win out in diachrony, it makes sense to require that the abundance x increase if and only if $f_1 > \varphi(x)$, that is, whenever the fitness of the first variant is greater than average. This yields the following simple ansatz for a difference equation that governs the evolution of x over discrete time, t:

(2) $$x(t + 1) = \frac{f_1}{\varphi(x(t))} x(t)$$

It is plain to see that the value of x increases if $f_1 > \varphi(x)$ and decreases if $f_1 < \varphi(x)$. Moreover, simple algebra shows that, for all $x \in (0, 1)$, $f_1 > \varphi(x)$ holds if and only if $f_1 > f_2$ (and $f_1 < \varphi(x)$ if and only if $f_1 < f_2$). Thus, in this case, the dynamics of x are completely determined by the relative magnitudes of the fitnesses f_1 and f_2 of the competing variants.

This is enough to predict whether the abundance x increases or decreases, but the simple equation actually tells us more, for it also gives the propagation process a particular *shape*. It is not difficult to solve the difference equation (2), whereupon one finds that

(3) $$x(t) = \frac{1}{1 + \left(\frac{f_2}{f_1}\right)^t \left(\frac{1}{x(0)} - 1\right)}$$

for a given initial condition $x(0)$. Plotting this equation with time t on the horizontal axis and the abundance $x(t)$ on the vertical axis reveals that the curve has a sigmoidal

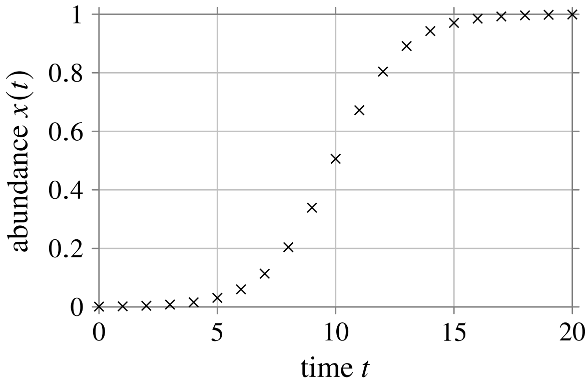

Figure 5. The solution (3) of the phenomenological equation (2) is a sigmoidal sequence of abundance values.

shape (Figure 5)—the shape of an S-curve so often mentioned in the literature (see Denison 2003)

Equation (2) provides a simple *phenomenological* model of the propagation of an innovative variant at the expense of a single competitor—phenomenological, as it reproduces some salient aspects of the phenomenon under study (especially, the sigmoidal nature of propagation) but is not derived from first principles (on this terminology, see Frigg and Hartmann 2020). The fitness quantities, so central to determining the outcome of propagation, remain nebulous: what are they exactly, and how could their values be empirically estimated? The next section provides one way of answering these questions by showing how the phenomenological equation can be derived from a model of language acquisition combined with a simple form of inter-generational transmission.

Even without a mechanistic foundation, the phenomenological model allows us to illustrate a few key dynamical notions. Notice that there are exactly three ways for equation (2) to satisfy the equilibrium condition $x(t+1) = x(t)$ in which no change occurs. These are $f_1 = \varphi(x(t))$, $x(t) = 0$ and $x(t) = 1$. Again, it is easy to check that $f_1 = \varphi(x(t))$ if and only if $f_1 = f_2$; in other words, a stationary state is predicted if no fitness difference obtains, no matter what the current abundance in the population. On the other hand, if $f_1 \neq f_2$, then only two equilibrium points exist: $x = 0$ and $x = 1$. This also makes intuitive sense: in a pure state, no variation exists and hence speakers will have no evidence of the existence of a competing variant which might be adopted instead of the conventional one.

When fitnesses are unequal, the two equilibrium states have different stabilities depending on whether $f_1 < f_2$ or $f_1 > f_2$. Suppose the former is the case. Then V_1 is less fit than V_2, and we would like our model to have the following property: starting from any point in the delta-dominance set $D_1(\delta)$ other than the pure state $x = 1$ in which V_1 has full abundance, we should expect to see the population state evolve towards the equilibrium $x = 0$ (full use of V_2) over time. Since $x(t+1) < x(t)$

Figure 6. Phase diagram of the phenomenological equation. The stable equilibrium is shown as a filled circle, the unstable one as an open circle. The stabilities reverse as the value of the fitness difference $f_1 - f_2$ crosses zero.

when $f_1 < f_2$, this is indeed what the phenomenological equation predicts; moreover, this behaviour occurs for any value of δ, i.e. for any amount of perturbation to the equilibrium state. The pure state $x = 0$ is a *globally stable* equilibrium in this case, while the state $x = 1$ is *unstable*. If $f_1 > f_2$, then we have what is essentially the mirror image of this situation, with $x = 1$ globally stable and $x = 0$ unstable (Figure 6).

5 Inter-generational evolution: deriving the phenomenological equation

The variational learning model (Yang 2002) provides an elegant account of language acquisition (see Unique ID WBCDL029). Utilizing a simple method to update the weights assigned to different grammatical options in response to environmental feedback (Bush and Mosteller 1955), the language learner tends towards a limiting distribution of those weights. In the two-variant case, the expected values of the weights (p_1, p_2) in the limit of an infinite learning period are

(4) $$\mathrm{E}[p_1] = \frac{c_2}{c_1 + c_2} \quad \text{and} \quad \mathrm{E}[p_2] = \frac{c_1}{c_1 + c_2}$$

where c_i denotes the penalty probability of variant i, the probability of the learner's environment punishing the use of this variant.

Now suppose that speakers are arranged in a sequence of non-overlapping generations in line with Andersen's (1973) classical Z-model: the input to the language learning process of generation $t + 1$ speakers comes from generation t speakers, who in turn received input from generation $t - 1$. With two competing variants, speakers may utter three kinds of strings: those which are only compatible with the first variant, those which are only compatible with the second variant, and those which are compatible with both. The latter make no difference to the dynamics, as they will not prompt the learner to update his or her grammar weights. The first two, however, play a central role. Let α denote the probability of a speaker of the first variant uttering a sentence not compatible with the second variant, and let β denote the probability of a speaker of the second variant uttering a sentence not compatible with the first variant (Figure 7). Then, if the abundance of the first variant in generation t is $x(t)$ (so that the abundance of the second variant is $1 - x(t)$), a generation $t + 1$ learner employing the first variant receives a penalty with a total

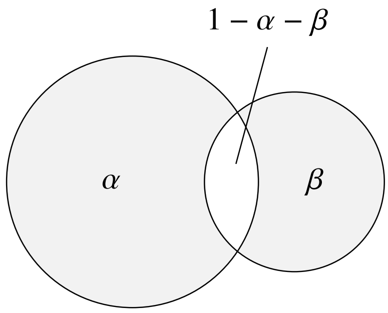

Figure 7. The probabilities of two competing variants generating output that is incompatible with the other variant constitute one possible source of differential fitness in language change. Here, the variant on the left generates some number of surface forms which the variant on the right does not (the grey set on the left). Similarly, the variant on the right has some number of such unique surface forms (the grey set on the right). The measures of these sets (the probabilities of drawing a surface form from each subset), α and β, determine the direction of inter-generational change, and thus play the role of (constant) fitness parameters.

probability of

(5) $$c_1 = \beta(1 - x(t)).$$

This is because that total probability decomposes into two independent contributions: first, the probability of encountering a speaker who may utter a string that clashes with the first variant in the first place $(1 - x(t))$, and secondly, the probability of this interlocutor in fact uttering such a string (β). Similarly a learner employing the second variant receives a penalty with a total probability of

(6) $$c_2 = \alpha x(t).$$

These observations allow us to write an inter-generational difference equation by substituting the above penalties in equation (4), whereby

(7) $$x(t+1) = \frac{\alpha x(t)}{\beta(1 - x(t)) + \alpha x(t)}$$

or in other words,

(8) $$x(t+1) = \frac{\alpha}{\varphi(x(t))} x(t)$$

where $\varphi(x) = \alpha x + \beta(1 - x)$.

This equation has the form (2); the probabilities α and β turn out to play the role of the fitness quantities f_1 and f_2. It is no longer mysterious where the fitnesses arise from: they simply reflect the probabilities with which incompatible strings are uttered by speakers of the corresponding variants (in a syntactic application, for instance, they may reflect the weak generative capacities of the competing grammars; see Unique ID WBCDL029 for extended discussion).

Deriving the phenomenological equation from a model of language learning thus gives us a *mechanistic* model (Craver 2006; Lindsey 2001; Baker et al. 2018) of propagation. The fact that the fitness quantities now receive an interpretation makes available a number of further tests and predictions. For instance, it is possible to estimate probabilities such as α and β from corpora, as a number of studies have done (e.g. Danckaert 2017; Heycock and Wallenberg 2013; Simonenko, Crabbé, and Prévost 2019; Yang 2000). This places an important independent check on the model—it is not sufficient that the model reproduce the trajectory of an observed historical change, it must also do so for values of the fitness parameters which are consistent with otherwise obtained estimates.

A further illustration of how a mathematically formulated mechanistic model generates predictions is the following: taking the solution (3), setting $x(0) = \delta$ and $x(t) = 1 - \delta$ and solving for t, one generates an estimate of the time required for the language community to traverse from a state of delta-dominance by one variant to a state of delta-dominance by another (for some suitably chosen small $\delta > 0$). Specifically, that time is

(9) $$t = 2\frac{\log(1/\delta - 1)}{\log(f_1/f_2)}$$

(assuming $f_1 > f_2$), and, with the inter-generational interpretation, this time is expressed in units of generations. Thus, it is possible to arrive at estimates of the passage times of propagation processes and to compare them against empirical data obtained from corpora, thereby yielding a further check on the mathematical model (see Ingason, Legate, and Yang 2013).

Two further remarks are in order. First, it needs to be stressed that variational learning in an inter-generational sequence of learner populations is not the only model that derives the phenomenological equation (2). A simple model involving categorical learners, in which speakers switch between grammatical variants categorically instead of entertaining probabilistic grammar weights, predicts the same population-level equation (Niyogi 2006), and similar equations are derived as the "mean dynamic" (the expected motion) of a number of stochastic processes that model competition in general populations (McKane and Newman 2004; Sandholm 2010). This is a common feature of complex systems more generally: multiple lower-level mechanisms may give rise to the same higher-level phenomenon, meaning that the latter alone can never provide a "crucial experiment" (Popper 1972) for the former. Secondly, above I have glossed over a number of idealizations that must be made in order to derive the deterministic equation (7): these include the assumptions that learners have infinite time at their disposal and that each learner in a given generation accesses the exact same learning environment. These assumptions are obviously false; however, making them brings out the deterministic "core" of the model. It is then possible to ask how relaxing those assumptions will affect our predictions. These questions will be taken up again in Sections 7–8, where I discuss stochastic models and models with local effects.

6 Beyond the phenomenological equation, I: variable fitness landscapes

Another way in which the above model idealizes is through its use of fitness parameters f_1 and f_2 which are constant—they are dependent neither on the abundances of the two competing variants in the speech community nor on any conceivable external influences on the linguistic system which themselves may change over time. This is appropriate in some situations. For instance, when variational learning is applied to a syntactic case study, the constant fitness parameters have the relatively unproblematic interpretation of referring to the relative weak generative capacities of the competing grammars. In other situations, however, particularly when a sociolinguistic phenomenon is being modelled, the assumption of constant fitnesses is less appropriate, since it matters *who* happens to be uttering the sentences the learner receives. In such cases we would like to relax the constant-fitness assumption and write

(10) $f_i = f_i(\mathbf{x}, \alpha, \beta, \ldots, \rho, \sigma, \ldots)$

to emphasize that the fitness parameters are *functions* of the population state $\mathbf{x}$ (the vector of current abundances of all variants in the speech community); of formal, language-internal parameters $\alpha, \beta, \ldots$; as well as of a number of potential language-external parameters $\rho, \sigma, \ldots$ (these could include, for instance, demographic parameters such as population size, social network connectivity, the relative proportion of second-language speakers in the community, and so on). This move is conceptually very important: it highlights the fact that linguistic forms rarely have a fitness (a functional advantage, a processing benefit, etc.) in isolation from other factors. To take a simple example, it is well-known that typological (near-)universals lead to strong correlations between features within a language (Greenberg 1963); an equivalent way of phrasing this is to say that, for instance, prepositions have greater fitness than postpositions in the context of a VO language but lower fitness in the context of an OV language. Similarly, it has been suggested that complex morphology has a lower fitness when the fraction of those learning the language as a second language is higher in the community (Trudgill 2011; also see Unique ID WBCDL046), illustrating one way in which the fitness of linguistic forms may depend on extra-linguistic parameters.

Linguistic variants in general thus compete within a *fitness landscape* which is rarely flat. As soon as the abundances of the competing variants change, fitness is likely to change, too, either climbing uphill or rolling down. This leads us to models of propagation which allow frequency-dependent fitness, and to the realm of evolutionary game theory (EGT). Originally developed (in the 1970s and 1980s) to model competition and cooperation in animal communities (see e.g. Maynard Smith 1982), EGT has since become a versatile tool not only in the biological sciences but also in economics, sociology and indeed in various subfields of linguistics. Jäger (2007) provides an EGT account of the typology of case marking, Deo (2015) and Yanovich (2017) discuss an EGT model of semantics and pragmatics, Baumann and Ritt (2017) provide a detailed analysis of lexical stress in terms of EGT, Kauhanen (2020) uses nonlinear fitness functions to model one type of sociolinguistic stratification, and Michaud (2020) shows how the utterance selection model of Baxter et al. (2006) can be interpreted in EGT terms, to mention a few examples.

To illustrate how a simple but non-constant fitness landscape may affect the behaviour of dynamical systems, let us return to equation (7). The system governed by this equation has two equilibria. For any choice of the (constant) advantage parameters α and β, $x = 0$ and $x = 1$ are fixed points. The stabilities of these equilibria are, moreover, fully determined by the fitness difference $\alpha - \beta$. Although the model has some attractive properties, most notably the sigmoidal form of propagation, the topology of the variation space does have one seriously unrealistic aspect. Suppose $\alpha > \beta$, so that the state $x = 0$ is unstable. Suppose the speech community initially finds itself in this state, and that an innovation occurs which raises x to a non-zero but small value $x = \delta$. Since the equilibrium is unstable, the evolutionary consequence of this perturbation is that the perturbation will grow—the innovation will propagate, reaching the globally stable equilibrium $x = 1$ in the limit.

In other words, in this type of model, "[o]nce a grammar is on the rise, it is unstoppable" (Yang 2000, 239). But now recall Nettle's (1999) observation: innovative variants always face a threshold of propagation which they must overcome in order to be selected. The dynamics of our simple model are wholly at odds with this fact (assuming it is a fact) of language variation and change. An innovation, even if infinitesimally small, is guaranteed to propagate successfully as long as the fitness difference points in the right direction. This suggests that the model is missing an integral component, conformity (Burridge 2017): speakers should display some amount of attraction towards a delta-dominant variant even if a potential innovative competitor does enjoy a formal advantage.

Such a mechanism is not difficult to incorporate into our simple model. Instead of the constant fitnesses α and β, we introduce the following slightly more complex fitness functions which, crucially, are now dependent on the abundance x:

$$(11) \quad \begin{cases} f_1(x) = sx + (1 - s)\alpha \\ f_2(x) = s(1 - x) + (1 - s)\beta \end{cases}$$

As before, the fitnesses depend on a constant part which may again (though need not) be interpreted as referring to the competing grammars' weak generative capacities (the α and β terms). Additionally, each fitness function includes another term that is proportional to the variant's current abundance in the speech community (x for the first variant, $1 - x$ for its competitor). The new s parameter is a "slider" which can be varied (between 0 and 1) to adjust the relative importance of the two contributions to fitness.

Inclusion of the conformity term helps to safeguard the delta-dominance states against invasion by innovations, unless those innovations are sufficiently numerous. Suppose variant V_2 is the currently delta-dominant one but that V_1 enjoys a formal advantage, so that $x \approx 0$ but $\alpha > \beta$. If V_2 is dominant enough (x is close enough to 0), then the first term in the fitness function $f_1(x)$ is very small. Depending on the value of s, this may be enough to guarantee that the delta-dominant state with the

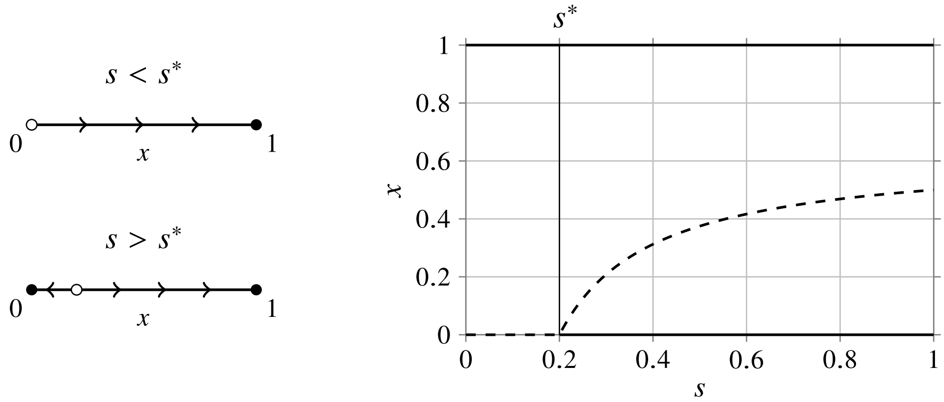

Figure 8. Including a mechanism for conformity in the simple model gives rise to a global invasion barrier (left bottom). This means that the equilibria $x = 0$ and $x = 1$ are locally stable for any choice of advantage parameters α and β, and therefore that an incoming innovation must cross an abundance threshold (the unstable equilibrium in between) in order to propagate. The invasion barrier exists if s, a parameter regulating the relative strength between conformity-driven fitness and non-conformity-driven fitness, has a high enough value. If it exceeds the critical value s^*, the invasion barrier exists, as shown in the bifurcation diagram (right), in which dashed lines indicate unstable and connected lines stable equilibria.

conservative variant is in fact locally stable, and hence that minority innovations cannot invade. It is not difficult to show that the s parameter has a critical value s^* such that, for all $s > s^*$, the equilibrium is locally stable. The critical value is

$$(12) \quad s^* = \frac{|\alpha_1 - \alpha_2|}{1 + |\alpha_1 - \alpha_2|}$$

and when the invasion barrier exists, it is located at

$$(13) \quad x^* = \frac{1}{2} - \frac{1 - s}{2s}(\alpha_1 - \alpha_2).$$

The former combination of one globally unstable and one globally stable equilibrium is now replaced by three local equilibria: two stable (the pure states $x = 0$ and $x = 1$) and one unstable in between. The unstable equilibrium represents the threshold in the sense of Nettle's (1999) threshold problem: any innovation must cross this *invasion barrier* in order to propagate. Another way of phrasing the same thing is to say that the conformity term induces a *basin of attraction* around the new, locally stable equilibrium $x = 0$; to succeed, an innovation must somehow be able to escape from this basin (Figure 8).

In the following section, we will see how this may happen through a combination of deterministic and stochastic factors.

7 Beyond the phenomenological equation, II: stochastic dynamics

The above results involve *deterministic* systems: in such systems, the current state $\mathbf{x}(t)$ completely determines all future states $\mathbf{x}(t')$ with $t' > t$. In an ideal world, perhaps, we might wish to maintain that an ultimate theory of language variation and change would have precisely this deterministic character, so that once all required information about initial conditions is to hand, all future states of the language community under consideration could be predicted deductive–nomologically. In practice, few scholars believe this level of theory can ever be attained. The factors affecting the trajectories of language communities are simply too numerous—the systems too complex—for deterministic prediction to be a realistic aim.

For these reasons, it makes sense to introduce a stochastic component into our models which is intended to stand for secondary aspects of the phenomenon under study. For instance, suppose we are interested in explaining syntactic change, and that a reasonably good, theory-driven picture of the linguistics involved is available. This picture will make reference to key terms such as ‘syntactic parameter’, ‘linguistic representation’, ‘generative capacity’, and the like, and the processes and entities picked up by these terms may reasonably be thought to be universal: they recur from one speaker to the next. It makes sense for such terms to fall under deterministic control in our models, and such is the case, for instance, with the constant fitness parameters α and β above, related in interpretation to the weak generative capacities of the competing grammars. As we have seen, these parameters can have a decisive influence on the dynamics of the speech community. On the other hand, it is intuitively clear that other factors play a role, too, such as speakers’ interaction patterns. However, the latter can hardly be modelled as deterministic factors: they are too numerous and complex (the resulting deterministic model, even assuming it were humanly possibly to write one down, would be so complicated as to hardly yield any insight). What is more, these factors cannot be experimentally controlled, at least not easily. It makes sense, then, to subsume such factors under a stochastic component of the formal model.

In the variational learning model (Section 5), the learner samples a variant to use and an input token to parse—both of these are, technically, random experiments. The learner’s trajectory is thus a stochastic process, and consequently, so is the trajectory of the speech community. Making suitable idealizing assumptions takes these models to their deterministic limits, yielding the deterministic equations explored above. Analysis of these equations can provide important insights, as we have seen; however, it also leaves important questions open. For example, above we saw that including a conformity term in the model sets up an invasion barrier for the innovative variant. In a deterministic model, Nettle’s threshold problem arises with full force: namely, if the innovation’s abundance is below the invasion barrier, then nothing can increase that abundance beyond that barrier. On the other hand, if the innovation’s abundance is above the barrier, then it is not really an innovation—its frequency in the speech community is already high enough to guarantee successful propagation.

So how is the invasion barrier—the threshold of propagation—crossed? If our

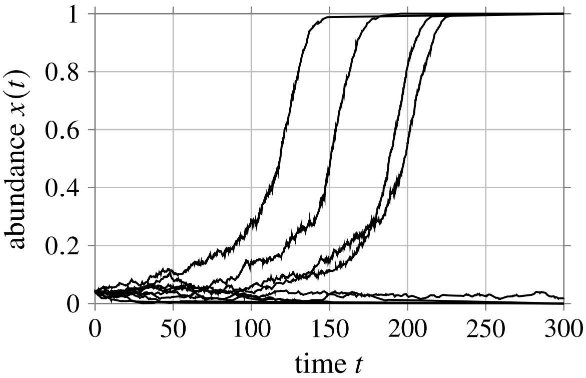

Figure 9. In a stochastic model, random fluctuations may help an initially small innovation cross the invasion barrier. Here, 10 realizations of the stochastic process were simulated, starting from an initial value $x(0)$ below the invasion barrier. In 6 of the 10 realizations, the innovation fails to pass the invasion barrier and dies out. In the remaining 4 realizations, it crosses the barrier and from there propagates successfully along an S-shaped trajectory. Time scaled by population size.

model provides a stochastic process instead of a deterministic trajectory, then it is possible for the speech community's state to "jump" unpredictably; in particular, it becomes possible for the invasion barrier to be crossed (for instance, due to the unpredictable nature of speaker interactions alluded to above). These jumps are not *entirely* unpredictable, nor entirely haphazard, however—our model need not have the character of a random walk or a purely neutral process (Kimura 1983). It is far more likely that many real-world processes of language variation and change involve a deterministic core, around which the stochastic process dances in a limited way. It is this interplay of deterministic and stochastic factors which gives us a full picture of the ensuing population dynamics.

To illustrate these concepts, let us return to the conformity-enhanced model from Section 6. Figure 9 displays a number of trajectories of the full stochastic process, simulated on the computer with the following assumptions. The advantage parameters were set at $\alpha = 0.2$ and $\beta = 0.1$; the conformity parameter had the value $s = 0.1$; the population consisted of $N = 1000$ individuals with 960 individuals speaking V_2 and 40 speaking V_1 at the start of the simulation, yielding an initial abundance of $x(0) = 40/1000 = 0.04$ for V_1. The particular choices $\alpha = 0.2$ and $\beta = 0.1$ yield $s^* = 0.09$, and since $s = 0.1 > 0.09 = s^*$, the invasion barrier exists; its value is $x^* = 0.05$, just slightly above the initial abundance $x(0) = 0.04$ (see equations 12–13).

Since $\alpha > \beta$, variant V_1 has a formal advantage over the initially delta-dominant variant V_2 and thus ought to replace it. However, the invasion barrier must be crossed. Looking at the simulated trajectories, we see that many realizations of the stochastic process simply fail: the innovation never manages to cross the threshold of propagation before dying away. These realizations conform to the predictions of the deterministic model: an innovation that starts from below the invasion barrier cannot propagate. A number of realizations, however, do cross the threshold. From there, the innovative variant is carried to delta-dominance with the help of the formal advantage difference. Even though the randomness inherent in the stochastic process means that the trajectory sometimes reverses direction, the trend given by the deterministic fitness difference ultimately means that most of the threshold-crossing trajectories carry the innovation to complete dominance.

It is also possible for further processes to be at play that facilitate an innovation's crossing the threshold of propagation. One such potential mechanism is *momentum selection*: under this hypothesis, speakers entertain an estimate of the direction of change of variables, and give more than proportional weight to variants which are estimated to be on the rise. For details, see Stadler et al. (2016).

These simple simulations illustrate the importance of bringing together deterministic and stochastic factors in formal models of language variation and change. For some types of models, it is possible to make substantial analytical progress in characterizing the distribution of the stochastically evolving quantities of interest—chiefly, how their expected values and variances change as functions of time. Classically, such an approach allows one to perform analytical calculations of key summary statistics such as the probability of a change occurring or the expected time until dominance. For an application of this logic of inquiry to an empirical case study, see Baxter et al. (2009).

8 Beyond the phenomenological equation, III: space and networks

Propagation of linguistic change does not occur in a vacuum, even though the foregoing discussion may have given this impression—rather, it occurs simultaneously across physical space and time. It may also propagate across different layers of society—a form of non-geographical propagation. Ultimately, over very long timescales, these processes of propagation conspire to give rise to both areal effects and genealogies (cf. Unique ID WBCDL055, Unique ID WBCDL056, Unique ID WBCDL060). To arrive at a more complete picture of language change, we thus need to look at models which incorporate such dimensions.

Some of the simplest models which can account for either geographical or sociological structure could be termed *two-compartment models*. In these models, two language communities (or subcommunities of a single community) are modelled, each having its own variation space and associated abundance vector. In addition to specifying how the dynamics of variants operate within each compartment, the model also specifies how variants "flow" from one compartment to the other. Such a model may be studied as a simple representation of language contact (the compartments are geographically contiguous areas) or of the diffusion of variants from one social class or age group to another (the compartments are sociologically defined). For examples of this sort of approach, see Mitchener (2011) and Kauhanen (2020, 2022).

It is straightforward, conceptually speaking, to generalize from such models to more general models defined on networks: each node of a network represents one compartment, and the links between nodes specify contact relations between compartments. At one extreme, the compartments may represent entire language communities; at the other, they may represent individual speakers or perhaps even parts of speakers (think of models of bilingualism or multilingualism). In general, the connectivity of the network may be arbitrarily complex: the degree distribution (the distribution of the number of connections) may be highly skewed, there may be multiple connections (of different types) between one and the same pair of nodes, and different connections may have different weights (representing, for instance, the frequency of interaction of the corresponding pair of nodes). Connections may also be directed in order to model asymmetries of communication or influence. On such approaches, see Ke, Gong, and Wang (2008), Fagyal et al. (2010), Kauhanen (2017), and Josserand et al. (2021).

In a system intended to model geographical diffusion, the nodes of the network are laid out in a special, typically regular and repeating fashion, for instance as a lattice. One key insight to arise from work on such models relates to the emergence and maintenance of extended geographical domains due to local influence. If all nodes on a geographical substrate (e.g. a lattice) evolved independently, we would observe a patchwork quilt of language states across that substrate, with little order in the spatial dimension (Figure 10, left). The emergence of extended spatial domains (connected components on the lattice; Figure 10, middle and right) requires physically nearby nodes to be correlated with each other—but how do such correlations arise? What is the *dynamic* process that gives rise to the *synchronic* geospatial ordering of linguistic variants? How do *global* properties of systems (such as language communities) emerge out of *local* processes between their parts (such as individual speakers)?

Above, it was pointed out that the S-shaped trajectories typical of the propagation of linguistic innovations may arise from various different underlying dynamic processes. The same holds true of the emergence of spatial order. Such order may arise from a simple copying dynamics, in which each node periodically updates its state by copying the state of one of its nearest lattice neighbours. Such a copying process can be taken as a very simple model of the local spatial propagation through successful inter-generational transmission; in more complex models, it is possible to model the entire acquisition process in each node over extended time. If this copying process is strong enough relative to other processes occuring independently within (as opposed to across) nodes, extended spatial domains arise; moreover, the size of these domains may be used to make inverse inferences about the susceptibility of different linguistic variables to undergo change (Kauhanen et al. 2021). Central here is the quantitative study of the isogloss junctions present in the system, defined simply as pairs of connected nodes on the lattice which differ in their state.

In models based on the above sort of copying (known as *voter models* in the literature on interacting particle systems; see Liggett 1985), the probability of a node accepting a variant from its spatial neighbourhood is proportional to the frequency of that variant in that neighbourhood. In such a model, the isoglosses are said to be *noise-driven*. An empirical alternative is that nodes are more (than proportionally)

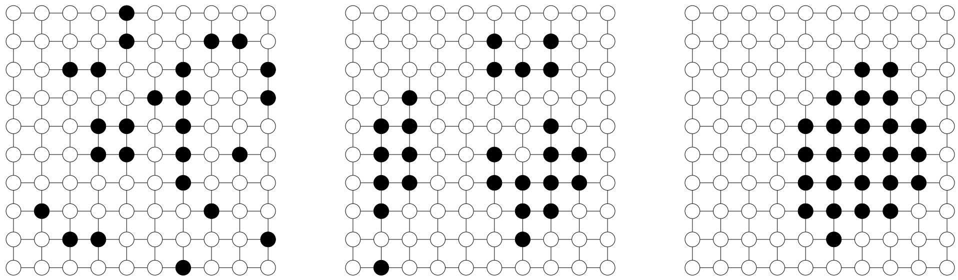

Figure 10. In this illustration, spatial order grows from left to right. Each panel shows a lattice in which each node is either black or white depending on which of two variants forms the majority option in that node. In each case, the number of black-majority nodes is 25 out of a total of 100 (10 by 10) nodes. A disordered system (left) has a large number of isogloss junctions, defined as connected node pairs which differ in their state. In a maximally ordered system (right), these junctions form a connected curve: an isogloss in the traditional dialectological sense. The emergence of spatial order out of dynamical processes is a central topic in the study of language dynamics, as it embodies both static properties of isogloss junctions (how numerous they are, what shape they take, where they are located) as well as dynamic ones (how they move, i.e. how innovations diffuse across space).

likely to adopt variants which are already locally popular. This leads to *conformity-driven* isoglosses (cf. Section 6). In a system driven by high conformity, isoglosses "prefer" to evolve to be maximally short, akin to how surface tension in a physical system such as a soap bubble produces the bubble's characteristic spherical shape, which minimizes surface area (Burridge 2018). Models of this kind, originally proposed in statistical physics as simple models of ferromagnets, are known as *Ising models* (Liggett 1985).

Burridge and Blaxter (2020) explored the question whether language dynamics are voter-like or Ising-like. Deriving voter and Ising dynamics from a single update rule in individual speakers in which the value of a conformity parameter can be varied to result in the two classes of model, and comparing the models' predictions about the resulting spatial domains against fieldwork data from the Survey of English Dialects (Orton and Dieth 1962), the authors found statistical support for the conformity-driven (Ising-like) update rule. Although the geospatial distributions of some features in the survey were consistent with either mechanism, in a majority of cases the conformity-driven model gave the better fit to data, suggesting that the cultural evolution of language does indeed include a conformity mechanism.

9 A note on technique: analytics, approximations, simulation

In most of the models discussed above, at least some progress can be made purely analytically: a pencil-and-paper analysis yields complete information about the possible range of behaviours of the model at least in some ideal limit (e.g. with the assumption of an infinite population or infinite learning time). This is not true of all models, of course. The more complex a model becomes, the less likely an analytical approach is to work. In such cases, computers can help: complex differential equations may be solved numerically, and discrete-time deterministic and stochastic models may be simulated step by step by implementing the model in some programming language.

Pioneering mathematical models of language variation and change, such as Niyogi and Berwick (1997), hail from a time in which the modest speed of personal computers still exerted significant challenges to the simulation of scientific models of any reasonable level of realism. Today's faster computers, particularly the possibility of parallel computing making use of multicore processors, grid computers or graphics processing units, make simulation of complex models in reasonable amounts of time a possibility.

This sort of modelling remains as much an art as a science, however, in the sense that it is important to use computational resources in sensible ways—the challenge being that very few explicit recipes exist to guide the researcher in what is sensible and what isn't. What can be said is that it generally makes little sense to write an overly complex simulation, trying to include in it as many aspects and processes of the real world as one can think of, only to be able to observe a limited number of simulation runs afterwards. A model that approaches in its complexity the phenomenon modelled becomes as opaque as the phenomenon itself; in particular, as the degrees of freedom of a model multiply, its dimensionality explodes and it becomes impossible to chart the model's potential behaviours systematically via simulation. A much more sensible approach is to begin with a simple, analytically tractable model, and drop one simplifying assumption at a time. This allows the researcher to explore what consequences making those simplifying assumptions had on the model's behaviours in the first place, to test whether the assumptions were really warranted. This strategy represents an epistemological sequence of models, starting from highly abstract, and often deterministic, macroscopic models, and ending at microscopic models defined at a lower level of analysis. Passage from the macroscopic to the microscopic domain may be facilitated by mesoscopic models which retain some characteristics of both extremes (Figure 11).

10 Conclusion: the three problems revisited

In Section 2, it was argued that any linguistic innovation must overcome three obstacles in order to eventually secure its place as the new convention: it must emerge in the first place (the actuation problem), it must turn from a minority innovation to a variant that stands the chance of propagating (the threshold problem), and it must have enough advantage, within the overall fitness landscape of the (socio)linguistic situation, to actually carry through to eventual delta-dominance (the sustain problem). The simple models we have discussed above can now be seen to shed light on each of these problems. It was argued that actuation itself is relatively unproblematic: linguistic mutations may either be purely random, like biological mutations, or they may be motivated by directionality. In either case, these

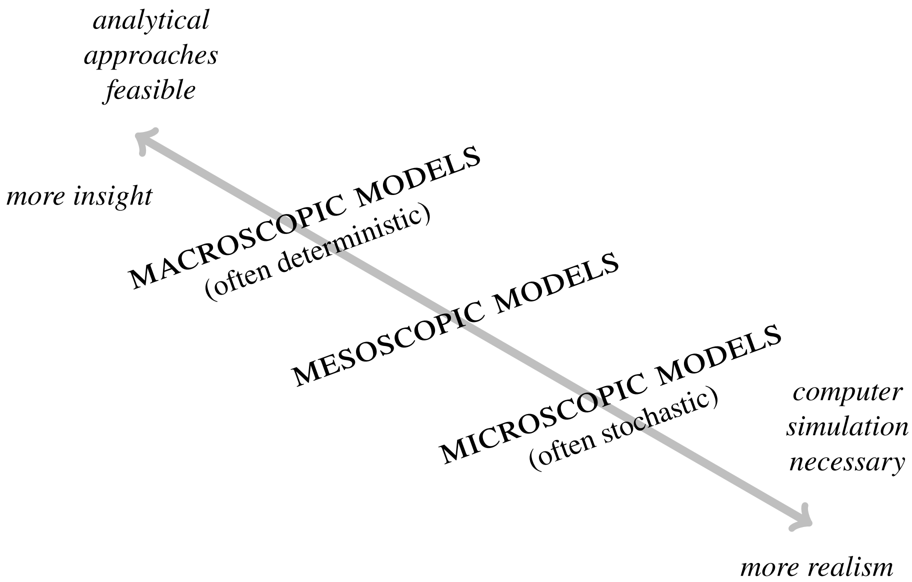

Figure 11. Computational models of language change can be roughly ordered on a scale from simple, macroscopic models to complex, microscopic models. The former allow analytical progress and have the potential to yield insight into the fundamental aspects of the phenomenon under study. By contrast, microscopic models, often stochastic in nature, offer more realism at the cost of analytical tractability. Mesoscopic models fall in between these extremes, and serve as an important link between the two.

mutations are constantly occurring in the production and parsing of language users. In general, such a mutation can stand the chance of propagating only if it manages to cross a threshold of propagation; the precise location of this threshold will vary from circumstance to circumstance, as it depends on complex interactions between linguistic and extra-linguistic parameters. In some cases, the threshold may be so low (and the inherent, formal linguistic advantage of the incoming variant so high) that most innovations succeed. In other cases, particularly if strong sociolinguistic motivations enforce high levels of conformity, very few innovations manage to cross the threshold. Actuation in this extended sense—innovation followed by escape from the basin of attraction of the conventional variant—must be seen as a stochastic process; it is the random fluctuations of finite populations which make crossing the threshold possible in the first place. Finally, the now viable innovation must have enough advantage for its propagation to be sustained until it becomes the only variant in use—or at least the majority one.

See also

WBCDL001
WBCDL029

WBCDL055
WBCDL116

Acknowledgements

Preparation of this manuscript has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 851423).

Data and code availability

The R (R Core Team 2021) code for replicating the simulations reported in Section 7 may be downloaded from https://github.com/hkauhanen/propagation.

References

Andersen, Henning. 1973. "Abductive and deductive change." *Language* 49 (4): 765–793.

Baker, Ruth E., Jose-Maria Peña, Jayaratnam Jayamohan, and Antoine Jérusalem. 2018. "Mechanistic models versus machine learning, a fight worth fighting for the biological community?" *Biology Letters* 14:20170660. https://doi.org/10.1098/rsbl.2017.0660.

Baumann, A., and N. Ritt. 2017. "On the replicator dynamics of lexical stress: Accounting for stress-pattern diversity in terms of evolutionary game theory." *Phonology* 34 (3): 439–471.

Baxter, Gareth J., R. A. Blythe, W. Croft, and A. J. McKane. 2006. "Utterance selection model of language change." *Physical Review E* 73:046118.

Baxter, Gareth J., Richard A. Blythe, William Croft, and Alan J. McKane. 2009. "Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English." *Language Variation and Change* 21 (2): 257–296.

Burridge, James. 2017. "Spatial evolution of human dialects." *Physical Review X* 7:031008. https://doi.org/10.1103/PhysRevX.7.031008.

———. 2018. "Unifying models of dialect spread and extinction using surface tension dynamics." *Royal Society Open Science* 5:171446. https://doi.org/10.1098/rsos.171446.

Burridge, James, and Tamsin Blaxter. 2020. "Using spatial patterns of English folk speech to infer the universality class of linguistic copying." *Physical Review Research* 2:043053. https://doi.org/10.1103/PhysRevResearch.2.043053.

Bush, Robert R., and Frederick Mosteller. 1955. *Stochastic models for learning.* New York, NY: Wiley.

Craver, Carl F. 2006. "When mechanistic models explain." *Synthese* 153:355–376. https://doi.org/10.1007/s11229-006-9097-x.

Danckaert, Lieven. 2017. "The loss of Latin OV: steps towards an analysis." In *Elements of comparative syntax: theory and description,* edited by Enoch Aboh, Eric Haeberli, Genoveva Puskás, and Manuela Schönenberger, 401–446. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781501504037-015.

Denison, David. 2003. "Log(ist)ic and simplistic S-curves." In *Motives for Language Change,* edited by Raymond Hickey, 54–70. Cambridge: Cambridge University Press.

Deo, Ashwini. 2015. "The semantic and pragmatic underpinnings of grammaticalization paths: the progressive to imperfective shift." *Semantics & Pragmatics* 8 (14): 1–52.

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. "Centers and peripheries: network roles in language change." *Lingua* 120:2061–2079.

Frigg, Roman, and Stephan Hartmann. 2020. "Models in Science." In *The Stanford Encyclopedia of Philosophy,* Spring 2020, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/.

Greenberg, Joseph H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements." In *Universals of human language,* edited by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.

Heycock, Caroline, and Joel Wallenberg. 2013. "How variational acquisition drives syntactic change: the loss of verb movement in Scandinavian." *Journal of Comparative Germanic Linguistics* 16:127–157.

Hopper, Paul J., and Elizaboth Closs Traugott. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Ingason, Anton Karl, Julie Anne Legate, and Charles Yang. 2013. "The evolutionary trajectory of the Icelandic New Passive." *University of Pennsylvania Working Papers in Linguistics* 19:91–100.

Jäger, Gerhard. 2007. "Evolutionary game theory and typology: a case study." *Language* 83 (1): 74–109.

Josserand, Mathilde, Marc Allassonnière-Tang, François Pellegrino, and Dan Dediu. 2021. "Interindividual variation refuses to go away: a Bayesian computer model of language change in communicative networks." *Frontiers in Psychology* 12:626118. https://doi.org/10.3389/fpsyg.2021.626118.

Kauhanen, Henri. 2017. "Neutral change." *Journal of Linguistics* 53 (2): 327–358.

Kauhanen, Henri. 2020. “Replicator–mutator dynamics of linguistic convergence and divergence.” *Royal Society Open Science* 7:201682. https://doi.org/10.1098/rsos.201682.

———. 2022. “A bifurcation threshold for contact-induced language change.” *Glossa: a journal of general linguistics* 7 (1): 1–32. https://doi.org/10.16995/glossa.8211.

Kauhanen, Henri, Deepthi Gopal, Tobias Galla, and Ricardo Bermúdez-Otero. 2021. “Geospatial distributions reflect temperatures of linguistic features.” *Science Advances* 7:eabe6540. https://doi.org/10.1126/sciadv.abe6540.

Ke, Jinyun, Tao Gong, and William S.-Y. Wang. 2008. “Language change and social networks.” *Communications in Computational Physics* 3 (4): 935–949.

Kimura, Motoo. 1983. *The neutral theory of molecular evolution.* Cambridge: Cambridge University Press.

Kirby, James, and Morgan Sonderegger. 2013. “A model of population dynamics applied to phonetic change.” In *Proceedings of the Annual Meeting of the Cognitive Science Society,* 35:776–781.

Lass, Roger. 1980. *On explaining language change.* Cambridge: Cambridge University Press.

———. 1997. *Historical linguistics and language change.* Cambridge: Cambridge University Press.

Liggett, Thomas M. 1985. *Interacting particle systems.* New York, NY: Springer.

Lindsey, J. K. 2001. *Nonlinear models in medical statistics.* Oxford: Oxford University Press.

Maynard Smith, John. 1982. *Evolution and the theory of games.* Cambridge: Cambridge University Press.

McKane, A. J., and T. J. Newman. 2004. “Stochastic models in population biology and their deterministic analogs.” *Physical Review E* 70 (4): 041902. https://doi.org/10.1103/PhysRevE.70.041902.

Michaud, Jérôme. 2020. “A Game Theoretic Perspective on the Utterance Selection Model for Language Change.” In *The Evolution of Language: Proceedings of the 13th International Conference (EvoLang13),* edited by A. Ravignani, C. Barbieri, M. Martins, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, K. Mudd, and T. Verhoef. https://doi.org/10.17617/2.3190925. http://brussels.evolang.org/proceedings/paper.html?nr=39.

Mitchener, W. Garrett. 2011. “A mathematical model of prediction-driven instability: how social structure can drive language change.” *Journal of Logic, Language and Information* 20:385–396. https://doi.org/10.1007/s10849-011-9136-y.

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

Niyogi, Partha. 2006. *The Computational Nature of Language Learning and Evolution.* Cambridge, MA: MIT Press.

Niyogi, Partha, and Robert C. Berwick. 1997. "A dynamical systems model for language change." *Complex Systems* 11:161–204.

Nowak, Martin A. 2006. *Evolutionary dynamics: exploring the equations of life.* Cambridge, MA: Belknap.

Orton, H., and E. Dieth, eds. 1962. *Survey of English Dialects.* Leeds: E. J. Arnold.

Popper, Karl R. 1972. *Conjectures and refutations: the growth of scientific knowledge.* Fourth, revised edition. London & Henley: Routledge & Kegan Paul.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Roberts, Ian, and Anders Holmberg. 2010. "Introduction: parameters in minimalist theory." In *Parametric variation: null subjects in minimalist theory,* edited by Theresa Biberauer, Anders Holmberg, Ian Roberts, and Michelle Sheehan, 1–57. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511770784.001.

Sandholm, William H. 2010. *Population games and evolutionary dynamics.* Cambridge, MA: MIT Press.

Simonenko, Alexandra, Benoit Crabbé, and Sophie Prévost. 2019. "Agreement syncretization and the loss of null subjects: quantificational models for Medieval French." *Language Variation and Change* 31 (3): 275–301. https://doi.org/10.1017/S0954394519000188.

Sóskuthy, Márton. 2015. "Understanding change through stability: a computational study of sound change actuation." 163:40–60. https://doi.org/10.1016/j.lingua.2015.05.010.

Stadler, Kevin, Richard A. Blythe, Kenny Smith, and Simon Kirby. 2016. "Momentum in language change: a model of self-actuating s-shaped curves." *Language Dynamics and Change* 6:171–198. https://doi.org/10.1163/22105832-00602005.

Trudgill, Peter. 2011. *Sociolinguistic typology: social determinants of linguistic complexity.* Oxford: Oxford University Press.

Walkden, George. 2017. "The actuation problem." In *The Cambridge handbook of historical syntax,* edited by Adam Ledgeway and Ian Roberts, 403–424. Cambridge: Cambridge University Press.

Weinreich, Uriel, William Labov, and Marvin I. Herzog. 1968. "Empirical foundations for a theory of language change." In *Directions for historical linguistics: a symposium,* edited by Winfred P. Lehmann and Yakov Malkiel, 95–195. Austin, TX: University of Texas Press.

Yang, Charles D. 2000. "Internal and external forces in language change." *Language Variation and Change* 12:231–250.

———. 2002. *Knowledge and learning in natural language.* Oxford: Oxford University Press.

Yanovich, Igor. 2017. "Analyzing imperfective games." *Semantics & Pragmatics* 10 (17): 1–23.