Spatial Evolution of Human Dialects

James Burridge[*]

Department of Mathematics, University of Portsmouth, Portsmouth PO1 3HF, United Kingdom

(Received 28 February 2017; revised manuscript received 9 May 2017; published 17 July 2017)

The geographical pattern of human dialects is a result of history. Here, we formulate a simple spatial model of language change which shows that the final result of this historical evolution may, to some extent, be predictable. The model shows that the boundaries of language dialect regions are controlled by a length minimizing effect analogous to surface tension, mediated by variations in population density which can induce curvature, and by the shape of coastline or similar borders. The predictability of dialect regions arises because these effects will drive many complex, randomized early states toward one of a smaller number of stable final configurations. The model is able to reproduce observations and predictions of dialectologists. These include dialect continua, isogloss bundling, fanning, the wavelike spread of dialect features from cities, and the impact of human movement on the number of dialects that an area can support. The model also provides an analytical form for Séguy's curve giving the relationship between geographical and linguistic distance, and a generalization of the curve to account for the presence of a population center. A simple modification allows us to analytically characterize the variation of language use by age in an area undergoing linguistic change.

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

I. INTRODUCTION

Over time, human societies develop systems of belief, languages, technology, and artistic forms that collectively may be called *culture*. The formation of culture requires individuals to have ideas, and then for others to copy them. Historically, most copying has required face-to-face interaction, and because most human beings tend to remain localized in geographical regions that are small in comparison to the world, human culture can take quite different forms in different places. One aspect of culture where geographical distribution has been studied in great detail is dialect [1].

In order to visualize the spatial extent of dialects, dialectologists have traditionally drawn *isoglosses*: lines enclosing the domain within which a particular linguistic feature (a word, a phoneme, or an element of syntax) is used. However, it is not usually the case that language use changes abruptly at an isogloss—typically there is a transition zone where a mixture of alternative features is used [1]. In fact, there is debate about whether the most appropriate way to view the geographical organization of dialects is as a set of distinct areas or as a continuum

without sharp boundaries [1–3]. Whereas an isogloss represents the extent of an individual feature, a recognizable dialect is typically a combination of many distinctive features [1,4]. We can attempt to distinguish dialects by superposing many different isoglosses, but often they do not coincide [3], leading to ambiguous conclusions.

The first steps toward an objective, quantitative analysis of the shapes of dialect areas were made by Séguy [5,6], who examined large aggregates of features, making comparison between lexical distances and geographic separations. Central to the quantitative study of dialects, called *dialectometry* (see Ref. [7] for a recent review), is the measurement of linguistic distance which, for example, can be viewed as the smallest number of insertions, deletions, or substitutions of language features needed to transform one segment of speech into another [8]. This "Levenshtein distance" was originally devised to measure the difference between sequences [9]. Using a metric of this kind, a set of dialect observations can be grouped into clusters according to their linguistic (as opposed to spatial) closeness [10–13]. The clusters then define geographical dialect areas.

The question we address is *why* dialect domains have particular spatial forms, and to give a quantitative answer requires a model. The question has been addressed in the past, famously (amongst dialectologists) by Trudgill and co-workers [1,14], with his "gravity model." According to this, the strength of linguistic interaction between two population centers is proportional to the product of their populations, divided by the square of the distance between them. The *influence* of a settlement (e.g., a city) *i* on

---

[*]james.burridge@port.ac.uk

another j is then defined to be the product of interaction strength with the ratio $P_i/(P_i + P_j)$, where P_i and P_j are the population sizes of settlements i and j. These additive influence scores may then be used to predict the progress of a linguistic change that originated in one city, by determining the settlements over which it exerts the greatest net influence. It is then predicted that the change progresses from settlement to settlement in a *cascade*. Predictions may also be made regarding the combined influence of cities on neighboring nonurban areas. The model has been partially successful in predicting observed sequences of linguistic change [1,15–17], and offers some qualitative insight into the most likely positions of isoglosses [14]. In this paper, we offer an alternative model, also based on population data, which makes use of ideas from statistical mechanics. Rather than starting with a postulate about the nature of interactions between population centers, we begin with assumptions about the interactions between speakers. From these assumptions about small-scale behavior we derive predictions about macroscopic behavior. This approach has the advantage of making clear the link between individual human interactions and population-level behavior. Moreover, we are able to unambiguously define the dynamics of the model and make precise predictions about the locations of isoglosses, the nature of transition regions between linguistic forms, and the most likely structure of dialect domains. There are links between our approach and agent-based models of language change [18], which directly simulate the behavior of individuals. The difference between this approach and ours lies in the fact that for us, assumptions about individual behavior lead to equations for language evolution which are macroscopic in character. These equations have considerable analytical tractability and offer a simple and intuitive picture of the large-scale spatial processes at play.

In seeking to model the spatial distribution of language beginning with the individual, we are encouraged by the fact that dialects are created through a vast number of complex interactions between millions of people. These people are analogous to atoms in the physical context, and when very large numbers of particles interact in physical systems, simple macroscopic laws often emerge. Despite the fact that dialects are the product of hundreds of years of linguistic and cultural evolution [4], and thus historical events must have played a role in creating their spatial distribution [19], the physical analogy suggests that it may be possible to formulate approximate statistical laws that play a powerful role in their spatial evolution.

A physical effect analogous to the formation of dialects is *phase ordering* [20]. This occurs, for example, in ferromagnetic materials, where each atom attempts to align itself with neighbors. If the material is two dimensional (a flat sheet), this leads to the formation of a patchwork of domains where all atoms are aligned with others in the same domain, but not with those in other domains. The boundaries between these regions of aligned atoms evolve so as to minimize boundary length [21,22]. The human agents who interact to form dialects behave in roughly the same way (as do some birds [23]). When people speak and listen to each other, they have a tendency to conform to the patterns of speech they hear others using, and therefore to "align" their dialects. Since people typically remain geographically localized in their everyday lives, they tend to align with those nearby. This local copying gives rise to dialects in the same way that short-range atomic interactions give rise to domains in ferromagnets. However, whereas the atoms in a ferromagnet are regularly spaced, human population density is variable. We show that as a result, stable boundaries between domains become curved lines.

While our interest is in the spatial distribution of linguistic forms, there are other properties of language for which parallels with the physical or natural world can be usefully drawn, and corresponding mathematical methods applied. For example, the rank-frequency distribution of word use, compiled from millions of books, takes the form of a double power law [24,25], which can be explained [24] using a novel form of the Yule process [26,27], first introduced to explain the distribution of the number of species in genera of flowering plants. Historical fluctuations in the relative frequency with which words are used have been shown to decay as a language ages and expands [25], analogous with the cooling effect produced by the expansion of a gas. Methods used to understand disorder in physical systems ("quenched" averages) have been applied to explain how a tendency to focus on topics controls fluctuations in the combined vocabulary of groups of texts [28]. A significant focus of current statistical physics research has been on the evolution and properties of networks [29], which have many diverse applications from the spread of ideas, fashions, and disease [30] to the vulnerability of the internet [31]. Real networks are often formed by "preferential attachment" where new connections are more often made to already well-connected nodes, leading to a "scale-free" (power-law) distribution of node degree. The popularity of words has been shown to evolve in the same way [32]; words used more in the past tend to be used more in the future. Beyond the study of word use and vocabulary, agent-based models such as the naming game [33], used to investigate the emergence of language, and the utterance selection model [34], used to model changes in language use over time, have been particularly influential. We follow the latter model by representing language use using a set of discrete *linguistic variables*. Spatial models motivated by concepts of statistical physics have also been used to study the spread of crime [35] and to devise optimal vaccination strategies to prevent disease [36]. The importance of the emergence of order in social contexts, and connections to statistical physics, may be found in a wide-ranging review [37] by Castellano *et al.*

II. SUMMARY FOR LINGUISTS

A. Contents of the paper

The aim of this paper is to adapt the theory of phase ordering to the study of dialects, and then to use this theory to explain aspects of their spatial structure. For those without a particular mathematical or quantitative inclination, the model can be simply explained: We assume that people come into linguistic contact predominantly with those who live within a typical travel radius of their home (around 10–20 km). If they live near a town or city, we assume that they experience more frequent interactions with people from the city than with those living outside it, simply because there are many more city dwellers with whom to interact. We represent dialects using a set of linguistic variables [1], and we suppose that speakers have a tendency to adapt their speech over time in order to conform to local conventions of language use. Our model is deliberately minimal: these are our *only* assumptions. We discover that, starting from any historical language state, these assumptions lead to the formation of spatial domains where particular linguistic variants are in common use, as in Fig. 2. We find that the isoglosses that bound these domains are driven away from population centers, that they tend to reduce in curvature over time, and that they are most stable when emerging perpendicular to borders of a linguistic domain. These theoretical principles of isogloss evolution are explained pictorially in Figs. 3, 4, and 5, and provide a theoretical explanation for a range of observed phenomena, such as the dialects of England (Fig. 7), the Rhenish fan (Fig. 10), the wavelike spread of language features from cities (Figs. 12 and 16), the fact that narrow regions often have "striped" dialects (Fig. 11), and that coastal indentations including rivers and estuaries often generate isogloss bundles. Our assumptions also lead to a mathematical expression for the relationship between linguistic and geographical distance—the Séguy curve—and a hypothesis regarding the question of when dialects should be viewed as a spatial continuum, as opposed to distinct areas (Fig. 19).

B. How might a linguist make use of this work?

Without using mathematics, but having understood our principles of isogloss evolution and considered the examples set out in this paper, further cases may be sought where the principles explain observations. If the principles cannot explain a particular situation or are violated, one might seek to understand what was missing from the underlying assumptions, or if they were wrong. Since the assumptions are so minimal, they cannot be the whole story, and a discussion of possible missing pieces is given in Sec. VIII. For the mathematically inclined linguist, Appendix A sets out an elementary scheme for solving the fundamental evolution equation on a computer. This scheme also offers a simple and intuitive understanding of the model, and can be implemented using only a spreadsheet (see Supplemental Material [38]), although a computer program would be much faster. Using this, isogloss evolution can be explored in linguistic domains with any shape and population distribution. The simplicity of the scheme invites adaptation to include more linguistic realism (e.g., bias toward a linguistic variant). Beyond the exploration of individual isoglosses, a line of inquiry that may be of interest to dialectometrists is to test our predicted forms of Séguy's curve against observations.

III. MODEL

Our aim is to define a model of speech copying which incorporates as few assumptions as possible, while allowing the effect of local linguistic interaction and movement to be investigated. The model has its roots in the ideas of the linguist Bloomfield [3], who believed that the speech pattern of an individual constantly evolved through his or her life via pairwise interaction. This microscopic view of language change led to the prediction that the diffusion of linguistic features should follow routes with the greatest density of communication. Bloomfield defined this as the density of conversational links between speakers accumulated over a given period of time. In our model, the analogy of this link density is an interaction kernel weighted by spatial variations in population distribution. We implicitly assume that interaction is inherently local so that linguistic changes spread via normal contact [39], rather than via major displacements, conquests, or dispersion of settled communities. We are, therefore, modeling language in stable settlements, with initial conditions set by the most recent major population upheaval.

We consider a population of speakers, each of whom has a small home neighborhood, and we introduce a population density $\rho(x, y)$ giving the spatial variation of the number homes per unit area. In order to incorporate local human movement within the model, we begin by defining a Gaussian interaction kernel for each speaker:

$$\phi(\Delta x, \Delta y) := \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{\Delta x^2 + \Delta y^2}{2\sigma^2}\right\}.$$

Note that the symbol $:=$ indicates the definition of a new quantity. Consider a speaker, Anna, whose home neighborhood is centered on (x_0, y_0). In the absence of variation in population density, ϕ is the normalized distribution of the relative positions, $(\Delta x, \Delta y)$, of the home neighborhoods of speakers with whom Anna regularly interacts. The constant σ, the interaction range, is a measure of the typical geographical distance between the neighborhoods of interacting speakers. Now suppose that density is not uniform due to the presence of a city or a sparsely populated mountainous area. In this case, while Anna is going about her daily life she is more likely to hold conversations with people whose homes lie in a nearby densely populated

region because these people constitute a greater proportion of the local population. To incorporate this density effect, we define a normalized weighted interaction kernel for a home at (x_0, y_0):

$$k(x_0, y_0; x, y) := \frac{\phi(x - x_0, y - y_0)\rho(x, y)}{\int_{\mathbb{R}^2} \phi(u - x_0, v - y_0)\rho(u, v)dudv}.$$

Given any region $\mathcal{A}$, the fraction of Anna's interactions that are with people who live in $\mathcal{A}$ is $\int_{\mathcal{A}} k(x_0, y_0, x, y)dxdy$.

We distinguish between dialects by constructing a set of linguistic variables whose values vary between dialects. A single variable might, for example, be the pronunciation of the vowel u in the words "but" and "up" [4]. In England, northerners use a long form, "boott" and "oopp," with phonetic symbol [ʊ], and southerners use a short version, [∧]. Considering a single variable which we suppose has $V > 1$ variants, we define $f_i(x, y, t)$ to be the relative frequency with which the ith variant of our variable is used by speakers in the neighborhood of (x, y), at time t. For mathematical simplicity, we assume that nearby speakers use language in a similar way, so that $f_i(x, y, t)$ varies smoothly with position.

People speak on average 16 000 words per day [40] and can take months or years (depending on their age and background) to adapt their speech to local forms [41,42]. Changing speech habits therefore involves a very large number of word exchanges, at least in the tens of thousands (comparable in magnitude to typical vocabulary size [43]). Although the rate at which individuals adapt their speech is not constant throughout life (it is particularly rapid in the young), adaptation has been observed even in late middle age [44]. To capture the cumulative effect of linguistic interaction we make use of a *forgetting curve*, which measures the relative importance of recent interactions to older ones. From a mathematical point of view, the simplest form for this curve is an exponential, and in fact there is some evidence from experiments involving word recall [45], which suggests that this is an appropriate choice. However, we emphasize that the curve, for us, is simply a way to capture the fact that current speech patterns depend on past interactions and that older interactions tend to be less important. With this in mind we make the following definition of the memory of a speaker from the neighborhood of (x, y), for the ith variant of a variable,

$$m_i(x, y, t)$$

$$:= \int_{-\infty}^{t} \frac{e^{(s-t)/\tau}}{\tau} \left[\int_{\mathbb{R}^2} k(x, y; u, v) f_i(u, v, s)dudv \right] ds \quad (1)$$

$$\approx \int_{-\infty}^{t} \frac{e^{(s-t)/\tau}}{\tau}$$

$$\times \left[f_i(x, y, s) + \frac{\sigma^2}{2\rho(x, y)} \nabla^2 \{\rho(x, y) f_i(x, y, s)\} \right] ds. \quad (2)$$

An intuitive understanding of this equation may be gained by imagining that each speaker possesses an internal tape recorder that records language use as they travel around the vicinity of their home. As time passes, older recordings fade in importance to the speaker, and the variable m_i measures the historical frequency with which variable i has been heard, accounting for the declining importance of older recordings. The rate of this decline is determined by the parameter τ, which we call *memory length*, and note that changing its value simply rescales the unit of time. We note also that this form of memory may be seen as a deterministic spatial version of the discrete stochastic memory used in the utterance selection model [34,46]. On the grounds that speakers collect very large samples of local linguistic information, our definition does not contain terms representing random sampling error. In going from Eq. (1) to (2) we use the saddle point method [47] to approximate the spatial integral in Eq. (1) and assume that $|\nabla^2\rho|/\rho$ is small compared to σ^2 (that is, population changes approximately linearly over the length scale of human interaction).

To allow speakers to base their current speech on what they have heard in the past, we let $f_i(x, y, t)$ be a function p_i of the set of memories $(m_1, m_2, ..., m_V) =: \mathbf{m}$:

$$f_i(x, y, t) := p_i[\mathbf{m}(x, y, t)].$$

Differentiating Eq. (2) with respect to t, and rescaling the units of time so that one time unit is equal to one memory length τ, we obtain

$$\frac{\partial m_i(x, y, t)}{\partial t} = p_i[\mathbf{m}(x, y, t)] - m_i(x, y, t)$$

$$+ \frac{\sigma^2}{2\rho(x, y)} \nabla^2 \{\rho(x, y) p_i[\mathbf{m}(x, y, t)]\}, \quad (3)$$

which governs the spatial evolution of the ith alternative for a single linguistic variable. We note that memory length no longer appears as a parameter. An enhanced intuitive understanding of this evolution equation may be gained from its discrete counterpart, used to find computational solutions, and derived in Appendix A.

The simplest possible choice for p_i is to let speakers use each variant with the same frequency that they remember it being used: $p_i[\mathbf{m}(x, y, t)] = m_i(x, y, t)$. This produces "neutral evolution" [34,46,48–50], where there is no bias in the evolution of each variant. Equation (3) then describes pure diffusion, and variants spread out uniformly over the system. If all linguistic variables evolved in this way we would eventually have one spatially homogeneous mixture of grammar, pronunciation, and vocabulary. If our memory model involved a stochastic component [46], then eventually we would expect all but one variant of each variable to disappear. Neither of these outcomes reflects the reality of locally distinctive forms of language.

We motivate our choice for p_i based on two observations. The first is that dialects exist. In order for this to be the case, if the ith variant of a linguistic variable has been established amongst a local population for a considerable time so that $m_i \approx 1$, then a small amount of immigration into the region by speakers using a different variable should not normally be sufficient to change it. Mathematically, this is equivalent to the statement that the nondiffusive term, $p_i[\mathbf{m}(x, y, t)] - m_i(x, y, t)$, in our evolution equation [Eq. (3)] must possess a locally stable fixed point at $m_i = 1$. The second observation is derived from experiments on social learning, which show that the behavior of individuals is considerably influenced by the majority opinion of those with whom they interact [51–53]. In fact, such social conformity is widely observed in the animal kingdom and is responsible for the formation of dialects in some species of birds [54]. Recent experimental research into human social learning [53], in which individuals were allowed to make a choice, before being exposed to the opinions of a group, has revealed that the likelihood of an individual switching their decision depends nonlinearly on the proportion of the group who disagree. It is an increasing function which climbs rapidly when the proportion exceeds 50%, also possessing an inflection for large groups. In the context of language, such nonlinear conforming behavior would mean that variants that were used more frequently than others should be used with disproportionately large frequency in the future. A simple way to capture this behavior is to define

$$p_i(\mathbf{m}) := \frac{(m_i)^\beta}{\sum_{j=1}^{V}(m_j)^\beta}, \tag{4}$$

where $\beta \geq 1$ measures the extent of conformity (non-neutrality). If $\beta = 1$, we have the neutral model, and for $\beta > 1$, the nondiffusive term in Eq. (3) has a stable fixed point at $m_i = 1$. According to Eq. (4), individuals disproportionately favor the most common variants they have heard: they have a tendency to *conform* to the local majority language use with β measuring the strength of this effect. In the limit $\beta \to \infty$, all speakers use only the most common (modal) dialect they have heard. An example of this function is plotted in Fig. 1 for the $V = 2$ model. Conforming behavior allows *local* dialects to form, as we show below.

The model we define is a coarse-grained description of real linguistic interactions which in reality are much more complex. Much of this complexity arises because there are often many distinct class, ethnic, or age-defined social networks in any given geographical region. Within each of these subgroups the need to conform leads to similar speech patterns among members, and these patterns often, but not always, spread to other groups. Research by linguists has demonstrated that social factors strongly influence the uptake of particular speech patterns [55] and that language use is correlated with social class and identity. In American
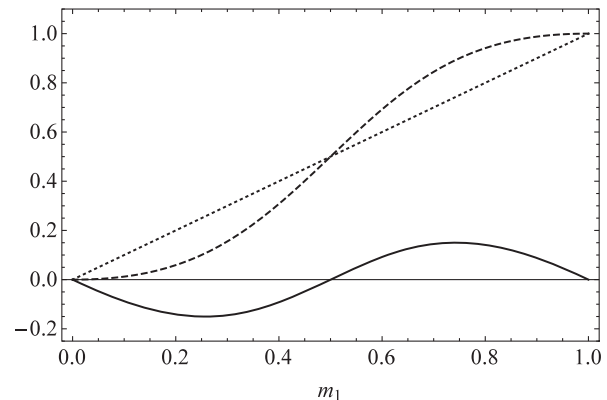

FIG. 1. Dashed line shows the function $p_1(\mathbf{m}) \equiv f_1$, defined by Eq. (4) for the $V = 2$ model with conformity number $\beta = 2$. Dotted line shows the neutral version $p_1(\mathbf{m}) = m_1$ (when $\beta = 1$) for comparison. Solid line shows the function $p_1(\mathbf{m}) - m_1$ in the case $\beta = 2$, giving the time derivative of the memory in the absence of spatial variation. Note that the dashed line shows that when speakers' memories contain a majority of variant 1, they use this variant in a greater proportion than they recall it being used. This leads to progressively greater levels of conformity: more speakers using variant 1.

English, for example, language change is often initiated by the working and lower middle classes [56,57], before spreading to other groups. Some forms of language change are driven by resistance to conformity; for example "prestige dialects" (received pronunciation in the UK) are used to signify membership of a social elite, set apart from the common people. The desire to set oneself apart from others can also create reversals in language use amongst subsets of a population. For example, local residents of Martha's Vineyard [58,59] reverted to an archaic form of pronunciation in order to reaffirm local tradition in the face of invading tourists. A similar effect was observed on the island of Ocracoke in North Carolina [60], but in this case the reversal was temporary. As well as social factors, language use may also be determined by age, gender, or ethnicity [61]. It is clear that reality is far more complex than our simple model, which does not make any of these distinctions between speakers. However, the fact that dialects exist is itself evidence that, in general, people do adapt to local speech patterns. To model every speaker as having the same need to conform is therefore a reasonable first approximation to reality. It also has the value of simplicity, allowing us later to determine the importance of various additional levels of complexity by comparing how effectively our model fits empirical data when compared to more complex models.

IV. SYNTHETIC DIALECT MAPS

A. Application to Great Britain

We apply our model to the island of Great Britain (GB), whose early inhabitants were known as Britons, and spoke

Celtic languages [62]. The earliest form of English was brought to the island by invading Germanic-speaking settlers. This became Anglo Saxon (or Old English), as written by Alfred, King of Wessex (849–899 A.D.), but would not be recognizable to modern speakers. It slowly changed, with external influences (notably Norman), into the English we know today [19].

We seek to discover the extent to which the spatial distribution of dialect structures that have emerged in GB can be predicted by Eq. (3). To model the evolution of individual linguistic variables we take mainland GB as our spatial domain, and numerically solve Eq. (3) on a grid of discrete points (Fig. 2) using an explicit Euler scheme [63] (Appendix A). The initial condition for the solution is a randomly generated spatial frequency distribution where each grid point is assigned a randomly selected variant. By repeatedly generating initial conditions and solving the system, we can determine the most probable equilibrium spatial distributions of language use. The population density $\rho(x, y)$ is estimated using 2011 census data [64], which gives the number of inhabitants at each of the $\approx 1.8 \times 10^6$ UK postcodes. A smooth density is then obtained from this by allowing the inhabitants to diffuse a short distance from the geographical center of their postcode. Despite significant overall population growth, the locations of major population centers in GB can trace their origins

back through hundreds of years. Since dialect evolution equation (3) depends only on relative population densities, the current density distribution therefore serves as reasonable proxy for historical versions. We estimate that σ lies in the range $5 < \sigma < 15$ km based on that fact that the average distance traveled to work in GB in 2011 was 15 km [64], whereas the average distance traveled to secondary school was 5.5 km [65]. In Sec, VII, we find that the typical width of a transition region between linguistic variables is $\approx 1.8\sigma(\beta - 1)^{-1/2}$. For example, the transition between northern and southern GB dialects is ≈ 60 km wide [1], which, if $\sigma = 10$ km, gives the approximation $\beta \approx 1.1$.

1. Evolution of isoglosses

When it comes to interpreting our results, the fact that usage frequencies are continuously varying through space presents a similar problem to that faced by dialectologists when trying to draw isoglosses. We resolve this by defining domain boundaries to be lines across which the *modal* (most common) variant changes. A domain is therefore a region throughout which a single variant is the most commonly used. We may think of domain boundaries as synthetic isoglosses generated by Eq. (3). In Fig. 2, we show a series of snapshots of the evolution of domains when there are $V = 3$ variants. Isogloss evolution is driven by a two-dimensional form of surface tension [66]: in the absence of density variation, curved boundaries straighten out. Figure 3 illustrates why this happens faster when curvature is greater. Here, speaker L hears more of variant A, so domain B will retract in this locality. Speaker R hears more of variant B, and so domain A will retract in this region. The net effect will be to straighten the boundary, reducing its length. If a boundary forms a closed curve, then this length reduction effect can cause it to evolve toward a circular shape, and reduce in area, eventually disappearing altogether. However, this *shrinking droplet* effect can be arrested or reversed if the droplet surrounds a sufficiently dense population center (a city). In fact, population centers typically repel isoglosses in our model, and so have a tendency to create their own domains. An explanation of this effect is given in Fig. 4. Here, we have a

FIG. 2. Evolution of the $V = 3$ model from randomized initial condition with $\sigma = 15$ km and $\beta = 1.1$ at times $t \in \{1, 2, 4, 8, 16, 32\}$, where one time unit corresponds to one memory length. Colors indicate which variant is most common at each position. Numerical solution implemented in C++ on grid with 2-km spacing [63] (GB is ≈ 1000 km north to south). Each grid point initialized with randomly selected variant.
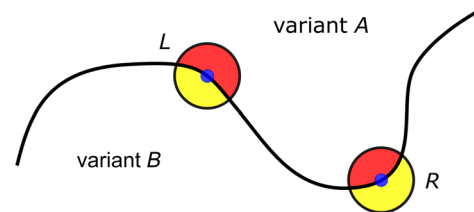

FIG. 3. The surface tension effect at domain boundaries. Blue dots represent speakers and black circles give an approximate representation of interaction ranges. In the red shaded parts of these interaction ranges, variant A is more common, and in the yellow shaded parts, variant B is more common.

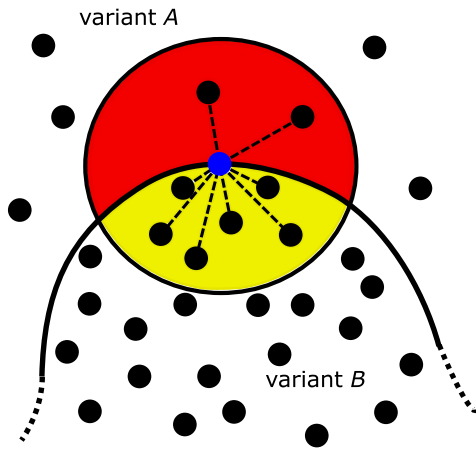FIG. 4. Behavior of an isogloss surrounding a densely populated area. The blue dot represents a speaker on the isogloss between variants A and B. Other speakers are shown as black dots. The circle around our speaker represents their typical range of interaction. Within red shaded part of this interaction range, variant A is more common, and in the yellow shaded part, variant B is more common. Because of variation in population density, they hear more of variant B (dashed lines indicate interactions) despite the fact that a greater *area* (red shaded) of their interaction range lies within the domain of variant A.
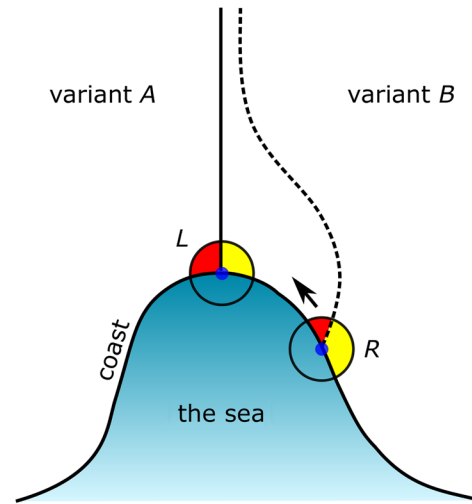

FIG. 5. Behavior of isoglosses at an indentation in coastline or other boundary (political or naturally occurring). Dashed isogloss is unstable and will evolve toward the solid isogloss which emerges perpendicular from the coast. We assume that both isoglosses are effectively anchored to a feature some distance away, opposite the boundary shown. Speakers are shown as blue dots, and colors have the same meanings as in Fig. 3.

region of high population density in which linguistic variant B is dominant, surrounded by a low-population density region where variant A is in common use. We consider the linguistic neighborhood of a speaker located on the isogloss separating the two domains. From Fig. 4 we see that although the majority of the speaker's interaction range lies in region A, she has many more interactions with those in region B, and is therefore likely to adapt her speech toward variant B, causing the isogloss to shift outward into the low-density area.

The most common form of stable isogloss generated by our model is a line, typically with some population density induced curvature, connecting two points on the boundary of the system. In order to be stable, such lines must emerge perpendicular from system boundaries, and as a result they are attracted to indentations in coastline, as illustrated in Fig. 5. In this figure, we consider two speakers located at the points where two possible isoglosses meet the coast (or other system boundary—a country border or a mountain range, for example). Speaker R, on the dashed isogloss, hears more of variant B because the isogloss is not perpendicular to the coast; it will therefore migrate upward toward the apex of the coastal indentation until it reaches the stable form shown by the solid line. This effect can be seen in Fig. 2, where the longest east-west isogloss has migrated so that it emerges from the largest indentation on the east coast of GB. In reality this indentation, called "the wash," is the site of an *isogloss bundle* (the coincidence of several isoglosses) separating northern [ʊ] from southern [ʌ] [67]. A similar bundle occurs at the largest indentation

in the Atlantic coast of France: the Gironde estuary [1], separating the *langue d'oc* from the *langue d'oil*. The fact that bundling at such locations is predicted by our model provides the first sign of the predictive power of the surface tension effect.

Having considered the evolution of a single linguistic variable, we now turn to modeling dialects. A dialect is typically defined by multiple linguistic characteristics, and we can capture this by combining many solutions to Eq. (3). In Fig. 6, we superpose the synthetic isoglosses for 20 binary ($V = 2$) linguistic variables. We see that there is a significant degree of *bundling* where many isoglosses follow similar routes across the system. Given that the initial conditions for each variable are distinct random frequency distributions (Fig. 2), these bundles represent highly probable isogloss positions: many different early spatial distributions lead to these at later stages of evolution. The key point here is that the final spatial structure of dialect domains is rather insensitive to the early history of the language: the effects of surface tension and population density draw many different isoglosses toward the same stable configurations. In this sense the surface tension effect is an "invisible hand" which, in the long term, can overpower historical population upheavals. However, we emphasize that our model predicts only the spatial structure of dialects and not their particular *sound*; this is very much determined by quirks of history and the initial state of the system. Figure 6 also illustrates the effect of human mobility on dialect structure. For a smaller interaction range (5 km), the structure of synthetic isogloss bundles is more complex, producing a larger number of distinct
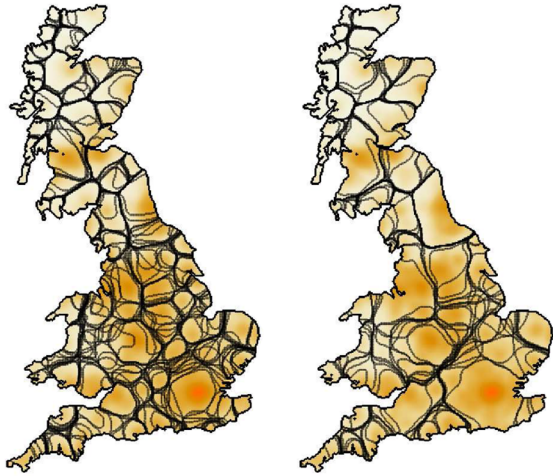
FIG. 6. Superposition of the isoglosses at $t = 50$ produced by 20 solutions of the $V = 2$ model with $\beta = 1.1$, each with different randomized initial conditions. For the left-hand map, $\sigma = 5$ km, and for the right-hand map, $\sigma = 10$ km (see video in Supplemental Material [38]). Background shading indicates population density with brightest orange corresponding to 7200 inhabitants per km^2.

regions. This effect is well documented in studies of the historical evolution of dialects which were, in the past, more numerous and covered smaller geographical areas [4]. Within our model, this is explained by the fact that fluctuations in population density become relevant to isogloss evolution only when they take place over a length scale that is comparable to the interaction range: Two human settlements could develop distinct dialects only if they were separated by a distance significantly greater than σ, otherwise they would be in regular linguistic contact.

2. Cluster analysis

Having analyzed our model using isoglosses, we now make comparison to recent work in dialectometry, where dialect domains have been determined using cluster analysis and by multidimensional scaling [68]. A typical clustering approach [10,12] is to construct a data set giving the frequencies of a wide range of variant pronunciations at different locations, and then to cluster these locations according to the similarity of their aggregated sets of characteristics. Resampling techniques such as bootstrap [69] may be used to generate "fictitious" data sets and improve stability. We mimic this approach by constructing a synthetic data set from 20 solutions of Eq. (3) with $V = 2$, and each with different random initial conditions, corresponding to different linguistic variables. We then randomly select a large number (6000) of sample locations within GB and determine the modal variants for each of the 20 variables at each location. This sample size is chosen to be sufficiently large so that the effect of resampling is only to make short length scale ($\ll 1$ km) changes to cluster
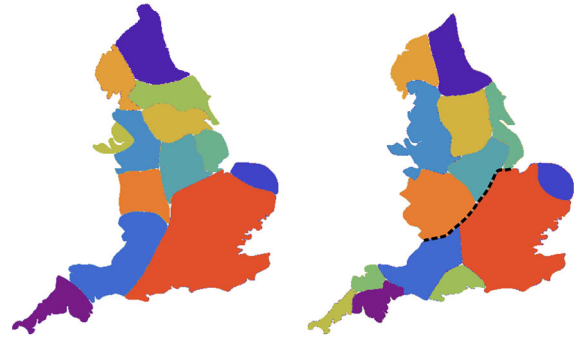

FIG. 7. Left map: Future England dialect boundaries predicted by Trudgill [4]. Right map: Future dialect boundaries predicted using k-medoids cluster analysis of 20 synthetic binary linguistic variables when $\sigma = 10$ km and $\beta = 1.1$ at $t = 150$. Levenshtein distance (or "edit distance") [9] used as distance metric. Colors, determined by Hungarian method, show mapping between dialect areas. Black dotted line shows north-south isogloss.

boundaries. These aggregated data are then divided into k clusters using the k-medoids algorithm [70] (available in the R language). The metric used for linguistic distance between sample points is the Manhattan distance between the binary vectors, where the two variants are labeled 1 or -1. Because we are comparing vectors which can be transformed into one another purely by substitutions (1 for -1 or vice versa), rather than insertions or deletions, this is equivalent to the Levenshtein distance used in dialectometry [2,9]. We find that almost identical results are obtained by applying Ward's hierarchical clustering algorithm [71] to the sample locations and subsequently cutting the tree into k clusters.

In order to compare our cluster analysis to the work of dialectologists, we consider a prediction for the future dialect areas of England (excluding Wales and Scotland) made by Trudgill [4], shown in the left-hand map of Figs. 7 and 8. This prediction divides the country into 13 regions, and is the result of a systematic analysis of regional variation in speech and ongoing changes. Such sharp

FIG. 8. Left map: Future England dialect boundaries predicted by Trudgill [4]. Right map: Voronoi tessellation with the same number of cells as Trudgill's prediction. Colors determined by Hungarian algorithm.

FIG. 9. Results of k-medoids cluster analysis of 20 synthetic binary linguistic variables when $\sigma = 10$ km and $\beta = 1.1$ at $t = 150$. Edit distance [9] used as distance metric. k values from left to right are $k = 16, 22, 30$.

TABLE I. Metrics measuring the similarity between Trudgill's predictions [4] for future English dialects and the predictions of our model. Metric acronyms are OL (overlap), WOL (weighted overlap), RI (Rand index), and ARI (adjusted Rand index). The Voronoi example column gives equivalent metrics for the example Voronoi tessellation in Fig. 8, and the Voronoi set column gives the mean metrics, with standard deviation, for five randomly generated Vornoi tessellations.

Metric	Model	Voronoi example	Voronoi set
OL	68%	41%	$45 \pm 3\%$
WOL	82%	36%	$49 \pm 11\%$
RI	0.91	0.84	0.83 ± 0.01
ARI	0.63	0.29	0.30 ± 0.01

divisions are a significant simplification of reality, however, and hide many subtle smaller-scale variations. The decision to define 13 regions therefore reflects a judgment on the range of language use which can be categorized as a single dialect. To allow comparison with this prediction, we perform a set of cluster analyses of near-equilibrium (large t) solutions for the whole of GB, for a range of values of the number k of clusters (see Fig. 9), with the aim of producing 13 within the subset of GB defined by England. The closest result is 14 clusters for $20 \leq k \leq 24$, with almost identical results within England for each of these choices. Having defined our synthetic dialect regions, we apply the Hungarian method [72] to find the mapping between our synthetic dialects and Trudgill's predicted dialects, which maximizes the total area of overlap between the two. The results are shown in Fig. 7. To provide a measure of the effectiveness of our model in matching Trudgill's predictions, we also define a *null* model, which divides the country into regions at random, independent of population distribution and without reference to any model of speaker interaction. There are a number of models that generate random tessellations of space [73], many of which are motivated by physical processes such as fracture or crack propagation. We exclude such physical assumption and thus opt for the Voronoi tessellation [73], based on the Poisson point process: the simplest of all random spatial processes. Our null model is then a Voronoi tessellation of England (Fig. 8) using 13 points selected uniformly at random from within its borders, with dialects labeled to most closely match Trudgill's map, using the Hungarian method.

Having generated our synthetic dialect maps, we now quantify the extent to which they match the predictions of Trudgill. The null model, because of its lack of modeling assumptions, will reveal the extent to which our model is "better than random" at matching these predictions. We offer four alternative metrics of similarity in Table I. The simplest metric is overlap (OL): the percentage of land area which is identified as belonging to the same dialect as

Trudgill's prediction. The weighted overlap (WOL) weights overlapping regions in proportion to their population density: it gives the probability that a randomly selected inhabitant will be assigned to the same dialect zone by both maps. From Table I we see that this probability is high (82%) for our model, but lower for the random Voronoi model. We suggest that this is a result of the fact that population centers tend to *repel* isoglosses and, therefore, often lie at the centers of dialect domains. We examine this repulsion effect in more detail below. The final two metrics are commonly used to compare clusterings. Consider a set S of elements (spatial locations for us) that has been partitioned into clusters (dialect areas) in two different ways. Let us call these two partitions X and Y. The Rand index (RI) [74] is defined as the probability that, given two randomly selected elements of S, the partitions X and Y will agree in their answer to the question: are both elements in the same cluster? A disadvantage with using this index to compare dialect maps is that the larger the number of regions in the maps, the more likely it is that two randomly selected spatial points will not lie in the same cluster in either map. The index therefore approaches 1 as the number of dialect areas grows. This problem may be countered by taking account of its expected value if X and Y were picked at random, subject to having the same number of clusters and cluster sizes as the originals [75]. The "adjusted Rand index" (ARI) is then defined as

$$\text{ARI} := \frac{\text{RI} - \text{expected index}}{1 - \text{expected index}}. \qquad (5)$$

The ARI $\in [-1, 1]$ measures the extent to which a clustering is a better match than random to some reference clustering and is used by dialectometrists [76] in preference to the original Rand index. For us, the reference clustering is Trudgill's predicted dialect map, and the Rand and adjusted Rand indices in Table I measure similarity to this reference.

The primary conclusion that may be drawn from the indices in Table I is that by all measures our model provides

a better match than the null model (indices all differ by at least 3 standard deviations, and typically many more). Of particular interest is the weighted overlap probability (WOL = 82%). Isoglosses are typically repelled by population centers, so tend to pass through regions of relatively low density. Because of this the WOL may be viewed as a measure of the effectiveness of the model at determining the centers of dialect regions and is less sensitive to small errors in isogloss construction, explaining its high value. It is important to realize also that Trudgill's predictions may themselves be imperfect.

We now make some qualitative comments. The dotted line in Fig. 7 shows the location of our model's most dense north-south isogloss bundle. This is coincident with what is described by Trudgill as "one of the most important isoglosses in England" [14] dividing those who have [ʊ] in butter from those who do not. In our model, the fact that this border lies where it does is a result of the surface tension effect which attracts many isoglosses towards the two coastal indentations at either end (see video in Supplemental Material [38]). The fact that many randomized initial boundary shapes evolve toward this configuration, and that the configuration is seen as important by dialectologists, supports the hypothesis that surface tension is an important driver of spatial language evolution. We also note that the western extremities of GB (Cornwall and northwest Scotland) support multiple synthetic dialects in our model, and we suggest that this is due to a heavily indented coastline and the fact that high aspect-ratio tongues of land are likely to be crossed by isoglosses; a fact predictable by analogy with continuum percolation [21]. The southwest peninsula has historically supported three dialects.

In future work, the model might be tested by comparing its predictions to well-researched dialect areas. On example is Netherlands. Here, dialectologists have performed a cluster analysis [68] based on Levenshtein distances between field observations of 360 dialect varieties (corresponding to 357 geographical locations), revealing 13 significant geographical groupings. The extent to which a model is consistent with these groupings, accounting for variability caused by finite sample size, could be tested by generating equivalent clusterings for multiple fictitious dialect samples of the same size.

B. Bundles, fans, stripes, and circular waves

We now illustrate a number of well-known features of dialect distributions which may be qualitatively reproduced by our model. We consider first the isogloss bundle reported by Bloomfield [3] separating "High German" from "Low German." The bundle emerged from the tip of an indentation of the Dutch-German speech area (bordered to the east by Slavic languages) and ran roughly east-west before separating approximately 40 km east of the river Rhine, and fanning out around cities such as
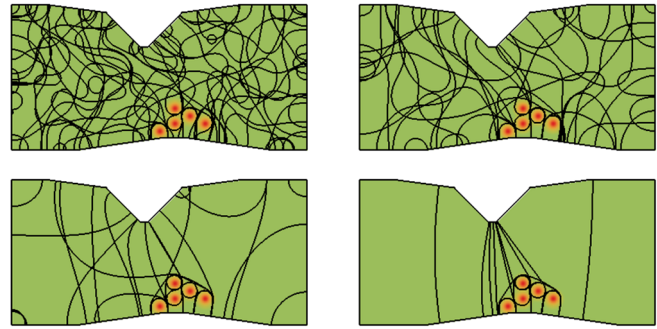

FIG. 10. Evolution of isoglosses in a 400×200 system with two opposing boundary indentations and unit background population density, together with a collection of cities contributing additional population densities $\rho(r) = (\rho_1 - 1) \exp\{-r^2/(2R^2)\}$, where r is distance from city center, $R = 10$, and $\rho_1 = 4$ (measuring ratio of peak city density to background). Parameter values $\sigma = 4$, $\beta = 2$. Evolution times $t = 10, 50, 300, 1660$. See video in Supplemental Material for full animation [38].

Dusseldorf, Cologne, Koblenz, and Trier. This arrangement of isoglosses is known as the "Rhenish fan." In Fig. 10, we construct an artificial system with boundaries approximating the geographical structure of relevant parts of the Dutch-German language area illustrated in Bloomfield [3] containing an artificial cluster of population centers representing the German cities located near the Rhine. The system was initialized using the same randomization procedure used for GB, and Fig. 10 shows a superposition of ten solutions, each with different initial conditions. In the early stages of evolution, very little pattern is discernible, but as time progresses, the main indentation collects isoglosses, while the cities repel them, producing a fanlike structure. We therefore suggest that the isogloss separation which created the Rhenish fan may have been the result of repulsion by the cities of the Rhine.

We next consider an example of what some physicists refer to as "stripe states" [21]: in finite systems that experience phase ordering, and have aspect ratios greater than 1, boundaries between two orderings often form across the system by the shortest route (in a rectangle, joining two long sides). A collection of such boundaries forms a striped pattern of phase orderings. Figure 11 illustrates this effect, produced by Eq. (3). Our model therefore predicts such striped dialect patterns in long thin countries, and a particularly striking example of the effect may be seen in the dialects of the Saami language [77]. The Saami people are indigenous to the Sámpi region (Lapland), which includes parts of Norway, Sweden, and Finland. Their Arctic homeland forms a curved strip with a length which is approximately 5 times its average width. The region is divided into ten language areas, and the boundaries of all but two of these take a near-direct route between the two long boundaries of the system, forming a distinctive striped pattern. Another example is the dialects of Japan,
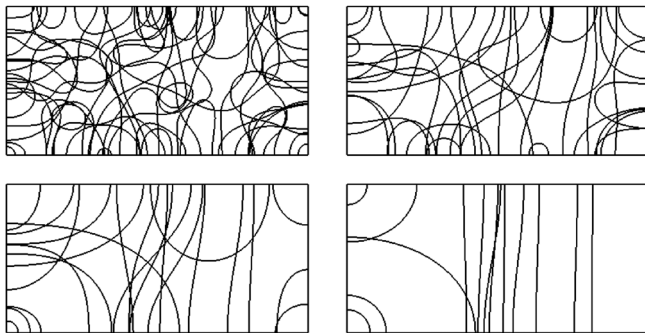
FIG. 11.   Evolution of isoglosses in a 400×200 rectangular system with uniform population density, starting from randomized initial conditions. We show a superposition of 10 such solutions. Notice that all isoglosses join the two long sides of the system. Parameter values $\sigma = 4$, $\beta = 2$. Evolution times $t = 30$, 90, 270, 810.

FIG. 12.   Evolution of a circular isogloss (initial radius 30) in a 200×200 system with unit background population density, together with a central city contributing additional density $\rho(r) = (\rho_1 - 1) \exp\{-r^2/(2R^2)\}$, where r is distance from city center, $R = 40$, and $\rho_1 = 21$. Parameter values $\sigma = 4$, $\beta = 2$. Evolution times $t \in \{10, 20, 30, \ldots, 300\}$.

whose boundaries in many cases cross the country perpendicular to its spine [78].

The relationship between geographical separation and linguistic distance (often measured using Levenshtein distances [2]) is typically sublinear [2,39]. The definition of linguistic distance and its relation to geographic distance was made by Séguy [5,6], and the relationship therefore goes by the name Séguy's curve [39]. It has been substantially refined and tested since [2,39,79], and also generalized to involve travel time [80]. Séguy's curve is not universal, however. For example, an analysis of Tuscan dialect data [81] reveals an unusually low correlation between phonetic and geographical distances. A more detailed analysis reveals that there are geographically remote areas which are linguistically similar, and that within an approximately circular region (radius ≈ 40 km) around the main city, Florence, phonetic variation correlates more strongly with geographical proximity. It is hypothesized [81] that this pattern marks the radial spread of a linguistic innovation (called "Tuscan-gorgia"). These Tuscan data motivate our final example of the qualitative behavior of our model. To illustrate how a linguistic variable can spread outwards from a population center, purely through the effects of population distribution and not necessarily driven by prestige or other forms of bias, we simulate our model using an artificial city with Gaussian population distribution (Fig. 12). The system is initialized with a circular isogloss, centered on the city, representing a local linguistic innovation. Because population density is a decreasing function of the distance from the city center, speakers on the isogloss hear more of the innovation than their current speech form, allowing it to expand (as explained in Fig, 4). We see in Sec. VII that this expansion will not not necessarily continue indefinitely. Expansion processes such as this have also been observed in Norway [14]. In that case, the progress of new linguistic forms was show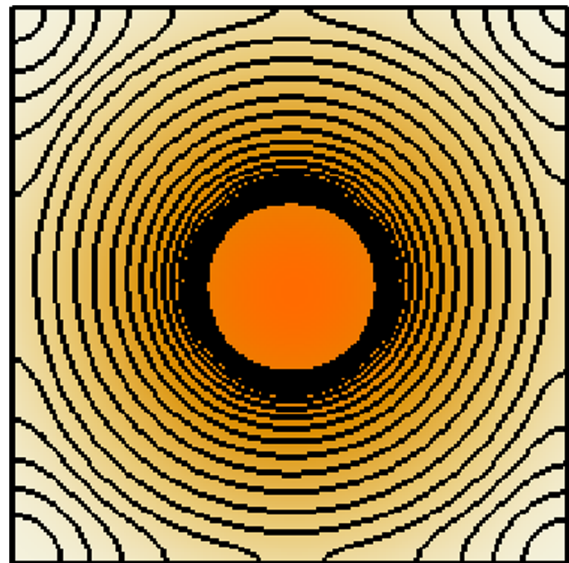n to depend on age, with changes more advanced for younger speakers, who are more susceptible to new forms of speech. We illustrate how this effect can be analyzed in Appendix B.

V. SÉGUY'S CURVE

We now determine the relationship between geographical and linguistic distance within our model, providing an analytical prediction for the form of Séguy's curve [5,6,39]. For simplicity, we consider the two variant model $V = 2$ and suppose that our language contains a number n of linguistic variables. At each location in space the local dialect is an n-dimensional vector of the local modal variants which we label 1 and -1. Letting $\phi(\mathbf{r}, t)$, where $\mathbf{r} = (x, y)$, be the vector field giving the distribution of these variants, the number of differences (the Levenshtein distance [9]) between two dialects $\phi(\mathbf{r}_1, t) =: \phi(1)$ and $\phi(\mathbf{r}_2, t) =: \phi(2)$ is $[n - \phi(1) \cdot \phi(2)]/2$. The linguistic distance $L(1, 2)$ between two dialects may be defined [82] as the fraction of variables that differ between them:

$$L(1, 2) := \frac{1}{2}\left(1 - \frac{\phi(1) \cdot \phi(2)}{n}\right). \tag{6}$$

Since we assume that each variant evolves independently of every other, the expected linguistic distance is

$$l(1, 2) := \mathbf{E}[L(1, 2)] = \frac{1}{2}(1 - \mathbf{E}[\phi_i(1)\phi_i(2)]), \tag{7}$$

where ϕ_i is the ith component of ϕ. To compute $\mathbf{E}[\phi_i(1)\phi_i(2)]$, we make use of the close similarity between

Eq. (3) and the time-dependent Ginzburg-Landau equation [20] to derive (see Appendix C) an analogue of the Allen-Cahn equation [83] giving the velocity of an isogloss at a point in terms of the unit vector $\hat{\mathbf{g}}$ normal to it at that point:

$$v = -\beta\sigma^2 \left[\frac{\nabla \cdot \hat{\mathbf{g}}}{2} + \frac{\nabla\rho \cdot \hat{\mathbf{g}}}{\rho} \right] \qquad (8)$$

$$= -\beta\sigma^2 \left[\frac{\kappa}{2} + \frac{\nabla\rho \cdot \hat{\mathbf{g}}}{\rho} \right]. \qquad (9)$$

The quantity κ is the curvature of the isogloss at the point: in the absence of variations of population density, the isogloss moves so as to reduce curvature. The second term in the square brackets produces a net migration of isoglosses towards regions of lower population density. To compute correlation functions between the field ϕ_i at different locations in space, we apply the Ohta-Jasnow-Kawasaki (OJK) method [84], introducing smoothly varying auxiliary field $m(x, y, t)$, which gives the value of the ith variant as $\phi_i = \text{sgn}(m)$. Note that the auxiliary field $m(x, y, t)$ is distinct from the memory $m_i(x, y, t)$ for the ith variant. The OJK equation, describing the time evolution of this field, adapted to include density effects, is (Appendix C)

$$\frac{\partial m}{\partial t} = \beta\sigma^2 \left(\frac{\nabla^2 m}{4} + \frac{\nabla\rho . \nabla m}{\rho} \right). \qquad (10)$$

We introduce the fundamental solution, $G(\mathbf{r}, t; \mathbf{r}_0)$ (the Green's function) of Eq. (10), giving the function $m(\mathbf{r}, t)$ subject to the initial condition $m(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)$. The solution for arbitrary initial conditions is then

$$m(\mathbf{r}, t) = \int_{\mathbb{R}^2} d\mathbf{r}_0 G(\mathbf{r}, t; \mathbf{r}_0) m(\mathbf{r}_0, 0). \qquad (11)$$

We assume that the initial condition of our system consists of spatially uncorrelated language use, so that $\mathbf{E}[\phi_i(\mathbf{r}_1, 0)\phi_i(\mathbf{r}_2, 0)] = \delta_{\mathbf{r}_1\mathbf{r}_2}$. A convenient, equivalent condition on the auxiliary field is to let it be Gaussian (normally) distributed $m(\mathbf{r}, 0) \sim \mathcal{N}(0, 1)$ with correlator

$$\mathbf{E}[m(\mathbf{r}_1, 0)m(\mathbf{r}_2, 0)] = \delta(\mathbf{r}_1 - \mathbf{r}_2). \qquad (12)$$

We can compute this correlator at later times using the fundamental solution G:

$$\mathbf{E}[m(\mathbf{r}_1, t)m(\mathbf{r}_2, t)]$$

$$= \int_{\mathbb{R}^4} G(\mathbf{r}_1, t; \mathbf{r}_1')G(\mathbf{r}_2, t; \mathbf{r}_2')\mathbf{E}[m(\mathbf{r}_1', 0)m(\mathbf{r}_2', 0)]d\mathbf{r}_1'd\mathbf{r}_2'$$

$$= \int_{\mathbb{R}^2} G(\mathbf{r}_1, t; \mathbf{r}_0)G(\mathbf{r}_2, t; \mathbf{r}_0)d\mathbf{r}_0. \qquad (13)$$

The linearity of our adapted OJK equation (10) ensures that $m(\mathbf{r}, t)$ remains Gaussian for all time [20] (to see this, note

that derivatives are limits of sums, and sums of Gaussian random variables are themselves Gaussian). However, the values of the field at different spatial locations develop correlations so that the joint distribution of any pair is bivariate normal. Following Bray [20], we define the normalized correlator

$$\gamma(\mathbf{r}_1, \mathbf{r}_2) := \frac{\mathbf{E}[m(\mathbf{r}_1, t)m(\mathbf{r}_2, t)]}{\sqrt{\mathbf{E}[m(\mathbf{r}_1, t)^2] \mathbf{E}[m(\mathbf{r}_2, t)^2]}}. \qquad (14)$$

Using the abbreviated notation $\gamma(\mathbf{r}_1, \mathbf{r}_2) \equiv \gamma(1, 2)$, the correlator for the original field may be found by averaging over the bivariate normal distribution of $m(\mathbf{r}_1, t) =: m(1)$ and $m(\mathbf{r}_2, t) =: m(2)$ (see, e.g., Ref. [85]):

$$\mathbf{E}[\phi_i(1)\phi_i(2)](t) = \mathbf{E}[\text{sgn}[m(1)]\text{sgn}[m(2)]] \qquad (15)$$

$$= \frac{2}{\pi}\sin^{-1}(\gamma(1, 2)). \qquad (16)$$

We now compute this correlator and derive a theoretical prediction for Séguy's curve.

A. Uniform population density

If population density is constant $\rho = C$, then our adapted OJK equation (10) reduces to OJK's original form, which has the fundamental solution

$$G(\mathbf{r}, t; \mathbf{r}_0) = \frac{\exp\{-\frac{|\mathbf{r} - \mathbf{r}_0|^2}{\beta\sigma^2 t}\}}{\pi\beta\sigma^2 t}, \qquad (17)$$

giving a normalized correlator

$$\gamma(1, 2) = \exp\left\{-\frac{r^2}{2t\beta\sigma^2}\right\} =: \gamma_t(r). \qquad (18)$$

Our prediction for Séguy's curve at time t is therefore

$$l(r, t) = \frac{1}{2}\left[1 - \frac{2}{\pi}\sin^{-1}(\gamma_t(r))\right]. \qquad (19)$$

This curve is plotted in Fig. 13 along with simulation results. We give the following interpretation of the curve. Starting from a randomized spatial distribution of language use, the need for conformity generates localized regions where particular linguistic variables are in common use, and these regions are bounded by isoglosses. These regions expand, driven by surface tension in isoglosses, so that from any given geographical point one would need to travel farther in order to find a change in language use. The linguistic distance between two points therefore tends to decrease with time, and the curve [Eq. (19)] gives the rate of decrease as exponential. There are features of reality which we might expect to alter this behavior. First, we assume that no major population mixing or migration takes place—such
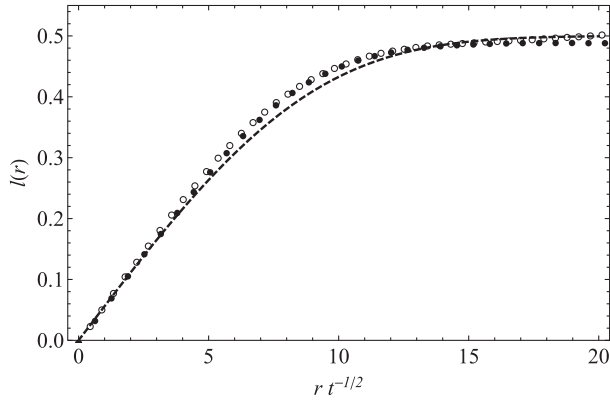
FIG. 13. Séguy's curve showing linguistic distance (l) versus geographical distance (r). Dashed line shows Eq. (19) in the case $\sigma = 4$, $\beta = 2$. Open and closed dots show simulated linguistic distances using same parameter values in a 1000×1000 system at times $t = 80, 160$. Note that linguistic distance depends only on the combination $rt^{-1/2}$, so curves evaluated at different times collapse onto one another.

events would have the effect of resetting the initial conditions of the model. Our prediction is valid only during times of stability. Second, we assume that the population is uniformly distributed in the system when in reality populations are clumped and, as we have seen, population centers can support their own dialects if they are large enough. We take some steps toward addressing this issue below. In Appendix D, we briefly discuss a simple one-dimensional simulation model from the dialectology literature [39], which includes the same large r behavior as in Eq. (19) for a particular choice (quadratic) of macroscopic "influence" curve.

B. Peaked population density

We now consider how Séguy's curve is modified by the presence of a peak in population density. In order to allow analytical tractability, we consider a simple exponentially decaying peak

$$\rho(x, y) = \exp\left\{-\frac{\sqrt{x^2 + y^2}}{R}\right\}, \qquad (20)$$

where $R > 1$. To understand the behavior of the modified OJK equation (10), it is useful to decompose it into an advection diffusion equation plus a source term:

$$\frac{\partial m}{\partial t} = \beta\sigma^2\left[\nabla \cdot \left(\frac{\nabla m}{4} + \frac{\nabla\rho}{\rho}m\right) - \left(\nabla \cdot \frac{\nabla\rho}{\rho}\right)m\right]. \quad (21)$$

Defining $r := \sqrt{x^2 + y^2}$, the average velocity field for the diffusing particle is

$$-\frac{\nabla\rho}{\rho} = \frac{(x, y)}{rR}. \qquad (22)$$

The source term is

$$-\left(\nabla \cdot \frac{\nabla\rho}{\rho}\right)m = \frac{m}{rR}. \qquad (23)$$

We now view Eq. (21) as describing the mass distribution for a collection of Brownian particles which are being driven radially away from the origin. The source term is interpreted as a field that causes particles to produce offspring at rate $(rR)^{-1}$ as they move through it. The fundamental solution, $G(\mathbf{r}, t; \mathbf{r}_0)$, to Eq. (21) is then the mass distribution for a very large (approaching infinite) collection of particles with total mass initially equal to one, all of which started at $\mathbf{r}_0$.

We wish to compute the dependence of linguistic distance on geographical distance from the peak of the population density (thought of as the center of a city). We therefore require the expectation

$$\mathbf{E}[m(\mathbf{0}, t)m(\mathbf{r}, t)] = \int_{\mathbb{R}^2} G(\mathbf{0}, t; \mathbf{r}_0)G(\mathbf{r}, t; \mathbf{r}_0)d\mathbf{r}_0. \quad (24)$$

Computation of a general closed-form expression for $G(\mathbf{r}, t; \mathbf{r}_0)$ is not our aim; preliminary computations in this direction suggest that if such a form existed, its complexity would restrict its use to numerical computations alone. Instead, we make arguments leading to a simple approximation for Séguy's curve. We observe first that the integrand in Eq. (24) is dominated by the region around $\mathbf{r}_0 = \mathbf{0}$. Numerical evidence for this is provided in Fig. 14, where we see that the fundamental solution grows in magnitude as $\mathbf{r}_0 \to \mathbf{0}$. In general, the solution consists of a circular plateau propagating outward from the origin plus an isolated but spreading peak also drifting away from the origin (Fig. 14). The plateau is formed once the rate of loss of particles from the peak source region ($|\mathbf{r}_0| \lesssim R^{-1}$) though
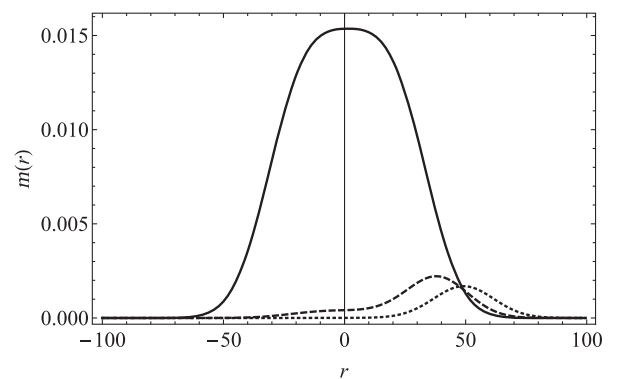

FIG. 14. Radial cross sections (along the line $y = 0$) through numerical approximations to fundamental solutions of the modified OJK equation (21) with population distribution Eq. (20) at time $t = 300$. Here, $r = \sqrt{x^2 + y^2} = x$. Parameter values are $\beta\sigma^2 = 1$ and $R = 10$. Initial conditions are $\mathbf{r}_0 = (1, 0)$, $(10, 0)$, $(20, 0)$ (solid, dashed, dotted lines, respectively).

advection and diffusion is equal to the rate of creation of new particles. The plateau height is determined by the particle mass which reaches the peak source region in the early stages of evolution. Because of radial drift, the only particles with a chance of doing this are those with sufficiently small Péclet number [86]:

$$\mathrm{Pe} = \frac{4|\mathbf{r}_0|}{R}, \tag{25}$$

where $\mathbf{r}_0$ is their starting point (or that of their earliest ancestor if they are daughters). Values of $\mathbf{r}_0$ that lie outside a region of radius $\propto R$ (henceforth Péclet region) around the origin can therefore be ignored when computing $G(\mathbf{0}, t; \mathbf{r}_0)$. For $R \gg 1$, the peak region forms a small fraction $O(R^{-4})$ of the Péclet region, and particles within the Péclet region have a probability of reaching the peak which decays exponentially with their initial distance from it. The function $G(\mathbf{0}, t; \mathbf{r}_0)$ will therefore itself be sharply peaked within the Péclet region, around $\mathbf{r}_0 = \mathbf{0}$, and we make the approximation $G(\mathbf{0}, t; \mathbf{r}_0) \approx h\delta(\mathbf{r}_0)$, where h is plateau height. Making use of this approximation in Eq. (24), we have

$$\mathbf{E}[m(\mathbf{0}, t)m(\mathbf{r}, t)] \approx hG(\mathbf{r}, t; \mathbf{0}). \tag{26}$$

To compute the variance

$$\mathbf{E}[m^2(\mathbf{r}, t)] = \int_{\mathbb{R}^2} G^2(\mathbf{r}, t; \mathbf{r}_0)d\mathbf{r}_0, \tag{27}$$

we note that if $|\mathbf{r}| \ll t/R$, then the dominant contribution to the integral comes from the plateau component of the solution. If $|\mathbf{r}| \gg t/R$, then the plateau will not have reached $\mathbf{r}$, so only the spreading peak component of the fundamental solution will contribute. Therefore,

$$\sqrt{\mathbf{E}[m^2(\mathbf{r}, t)]} \approx \begin{cases} O(1) & \text{if } |\mathbf{r}| \ll t/R \\ O(t^{-1/2}) & \text{if } |\mathbf{r}| \gg t/R. \end{cases} \tag{28}$$

We comment on the significance of this behavior below. To find the form of $G(\mathbf{r}, t, 0)$, we note first its circular symmetry, which reduces the number of variables in the OJK equation to two:

$$\frac{\partial m}{\partial t} = \frac{\sigma^2 \beta}{4}\left[\frac{\partial^2 m}{\partial r^2} + \left(\frac{1}{r} - \frac{4}{R}\right)\frac{\partial m}{\partial r}\right]. \tag{29}$$

We seek a traveling wave solution, subject to the initial condition $m(r, 0) = \delta(r)$, representing the expanding plateau, valid for large r, so that the $1/r$ term in Eq. (29) can be neglected. We obtain, as $t \to \infty$,

$$G(r, t; 0) \sim A\,\mathrm{erfc}\left(\frac{Rr - \beta\sigma^2(t - t_0)}{R\sigma\sqrt{\beta(t - t_0)}}\right), \tag{30}$$
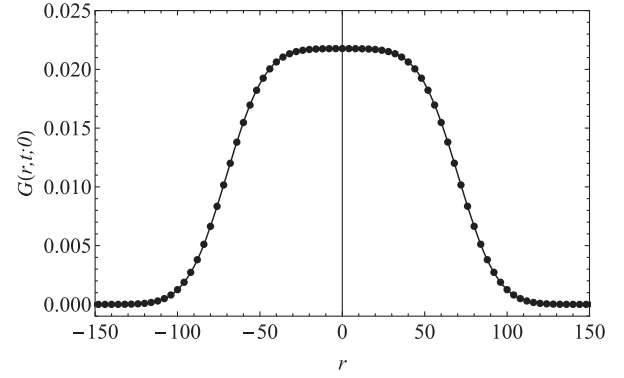

FIG. 15. Continuous line line shows radial cross section (along the line $y = 0$) through numerical solution of the modified OJK equation (21) with population distribution Eq. (20) at time $t = 700$, with initial condition $\mathbf{r}_0 = (0, 0)$. Here, $r = \sqrt{x^2 + y^2} = x$. Parameter values are $\beta\sigma^2 = 1$ and $R = 10$. Data points show asymptotic analytical solution $G(|r|, t; 0)$ [Eq. (30)], with $t = 700$, time offset $t_0 = -4.4$, and $A = 0.0109$ (found by maximum likelihood).

where t_0 is a time correction which accounts for the fact that the propagation velocity of the plateau takes some time to settle down to its long time value of $\beta\sigma^2/R$. We verify in Fig. 15 that this is the correct asymptotic solution by comparing it to the numerical solution of Eq. (29) for large t. We now approximate the normalized correlator as

$$\gamma(r, t) \approx \frac{G(r, t; 0)}{G(0, t; 0)}. \tag{31}$$

This approximation neglects the drop in the variance of $m(r, t)$ for $r \gg t/R$ described by Eq. (28), which amounts to neglecting a multiplicative factor $\sqrt{t}$ in the large r behavior of the correlator. Our approximate analytical prediction for Séguy's curve measured radially from the center of the exponentially decaying population distribution is therefore

$$l_c(r, t) := \frac{1}{2}\left[1 - \frac{2}{\pi}\sin^{-1}\left(\frac{G(r, t; 0)}{G(0, t; 0)}\right)\right]. \tag{32}$$

This prediction is compared to correlations in the full model (Fig. 16 and Fig. 17) by generating 100 realizations of isogloss evolution over the exponential population density, each with different randomized initial conditions. From Fig. 16 we see that as time progresses a growing region emerges around the center of the city in which the linguistic distance to the center is close to zero. An alternative visualization of this effect is given in Fig. 18, which shows a superposition of the isoglosses from 20 simulation runs. As time progresses, a circular patch emerges in the center of the system, which is devoid of isoglosses, and therefore where all speakers use the same linguistic variables. Outside of this central "city
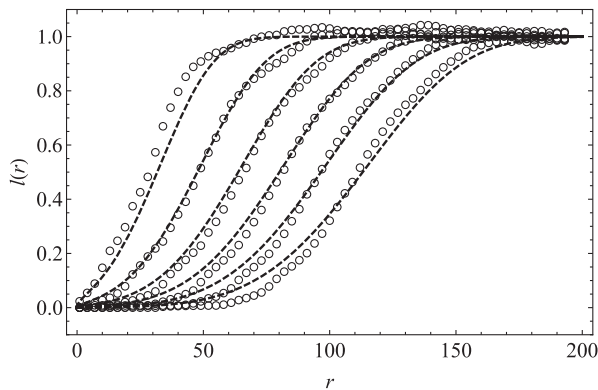
FIG. 16. Dashed lines show theoretical shape of Séguy's curve Eq. (32) centered at peak of population density $\rho = e^{-r/R}$ with $R = 20$. Curve computed using Eq. (32) when $\beta = 1.4$, $\sigma = 5$, times are $t = 10, 20, 30$ with offset $t_0 = -5.77$ (maximum likelihood estimate). Simulation points give equivalent correlations in the full model computed from 100 independent simulations in a 400×400 system.

FIG. 17. Dashed lines show theoretical shape of Séguy's curve centered at peak of population density $\rho = e^{-r/R}$ with $R = 20$, and with time evolution accelerated by factor of 1.25. Curve computed using Eq. (32) when $\beta = 1.1$, $\sigma = 5$, and times are $t_0 + 1.25t$, where $t = 10, 20, 30$ and $t_0 = 4.43$. Linear scaling of time determined by maximum likelihood fit of simulation to analytical prediction. Simulation points give equivalent correlations in the full model computed from 100 independent simulations in a 400×400 system at times $t = 10, 20, 30$.

dialect," we note that the asymptotic behavior of the complementary error function,

$$\text{erfc}(x) \sim \frac{\exp\{-x^2\}}{x} \quad \text{as } x \to \infty \tag{33}$$

together with the expansion $\sin^{-1}(\epsilon) = \epsilon + O(\epsilon^3)$ lead to the prediction that linguistic correlations fall as $e^{-c(\Delta r)^2}/(\Delta r)$, where c is a constant and Δr is the distance from the edge of the city dialect. This is a faster rate of decay than in the flat population density case. It appears from Fig. 16 that in reality the decay rate may be even faster than this. Further simulations reveal that the velocity with which the city dialect expands shows some systematic deviation from the prediction $v \approx \beta \sigma^2 / R$ of our OJK analysis. For example, in Fig. 17 we reduce the conformity parameter to $\beta = 1.1$ and we see that our theoretical predictions are in close agreement with the simulation data, provided we accelerate time by a factor of ≈ 1.25. The value of $\beta = 1.4$ selected in Fig. 13 produces a match between predicted and observed velocity, but for larger values of β the prediction is an overestimate. For example, when $\beta = 1.5$ with all other parameters identical, the simulated velocity in the full model is smaller than our prediction by a factor of 0.97. One possible explanation for this discrepancy is that the interface shape may affect the constant of proportionality in the Allen-Cahn equation (9), for example, if it did not match its constant density equilibrium form. We also note that OJK's assumption of isotropy in unit normals to isoglosses, although preserved globally by the circular symmetry of our system, is lost locally at the edge of the city dialect. Despite these shortcomings, the adapted OJK theory allows analytical
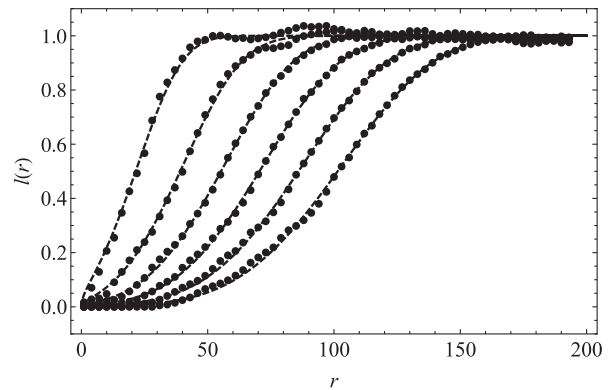
insight into the formation of dialects in population centers and the behavior of Séguy's curve around cities. We leave the development of a more sophisticated theory for future work.
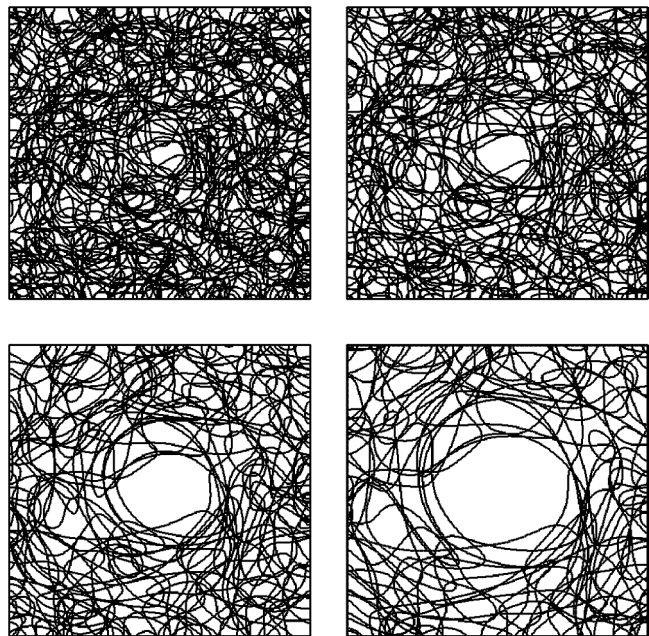

FIG. 18. Isogloss evolution in a 400×400 system with $V = 2$, $\beta = 1.1$, $\sigma = 5$ at $t = 10, 15, 25, 35$ with $\rho = e^{-r/R}$, where $R = 20$ and $r = 0$ corresponds to the center of the system. Plot is a superposition of 20 simulations with different initial conditions. Central peak repels isoglosses. See video in Supplemental Material for full animation [38].

VI. DIALECT AREAS AND DIALECT CONTINUA

There is debate amongst dialectologists as to the most appropriate way to view the geographical variation of language use [1,2]. The debate arises because it is rarely the case that dialects are perfectly divided into areas. Chambers and Trudgill [1] imagine the following example: we travel from village to village in a particular direction and notice linguistic differences (large or small) as we go. These differences accumulate so that eventually the local population are using a very different dialect from that of the village we set out from. Did we cross a border dividing the dialect area of the first village from that of the second, and if so, when? Alternatively, is it a mistake to think of dialects as organized into distinct areas; should we only think of a continuum?

We now set out what our model can tell us about these questions. In one sense, language use in our model is always continuous in space. Although domains emerge where one variable is dominant, domain boundaries form transition regions in which the variants change continuously (the width of these regions is computed in Sec. VII). Despite this, the boundary between two sufficiently large single-variant domains will appear narrow compared to the size of the domains, and in this sense the domains are well defined and noticeable by a traveler interested in one linguistic variable. Of more interest are the observations of a traveler who pays attention to the *full range* of language use. To perceive a dialect boundary, this traveler must see a major change in language use over a short distance. This change must be large in comparison to other, smaller changes perceived earlier. In our model a major language change is created by crossing a large number of isoglosses over a short distance. The question then is, under what circumstances will isoglosses bundle sufficiently strongly for dialect boundaries to be noticeable?

To answer this we need to recall the three effects which drive isogloss motion. First, surface tension, which tends to reduce curvature. Second, migration of isoglosses until they emerge perpendicular to a boundary such as the coast, the border of a linguistic region, a sparsely populated zone, or an estuary. Third, repulsion of isoglosses from densely populated areas. There are two major ways in which these effects can induce bundling, both of which require the essential ingredient of time and demographic stability in order for surface tension to take hold. Indented boundaries can collect multiple isoglosses, creating a bundle. Examples already noted include the Wash and the Severn in GB, the Gironde Estuary in France, and the historical indentation in the Dutch-German language area marking the eastern end of the Rhenish fan. A major boundary indentation may not always create a bundle, however: it may be that other parts of the boundary, or the presence of cities, creates a fanning effect. Variations in population density can also create bundling. Dense population centers which are large in comparison to the typical interaction range will push out linguistic change, and where two centers both repel, we expect to see bundling where their zones of influence meet. Each city would then create its own well-defined dialect area. Within real cities we also see subdialects spoken by particular social groups [55], but since our model does not account for social affiliations, we cannot explicitly model this.

In Fig. 19, we schematically illustrate examples of these effects using three imaginary "island nations." Nation *A* exhibits distinct dialect areas. The northernmost area is supported both by a city, which may generate and then repel language features, and by two indentations which form a "pinch point" which will *tend* to collect isoglosses via the boundary effect. Several other pinches exist which also collect isoglosses, creating distinct dialects. The southernmost city supports an isogloss via repulsion, which would otherwise migrate south under the combined influence of surface tension and the boundary effect, eventually disappearing. Nation *B* also possesses boundary indentations, but the lack of pinch points reduces bundling: while the indentations collect isoglosses, the smoother parts of the coastline allow isoglosses to attach anywhere, creating a continuum of language use. Two city dialects do exist, however, driven by repulsion. Finally, nation *C* is a convex region. This is an example of a system which, without a
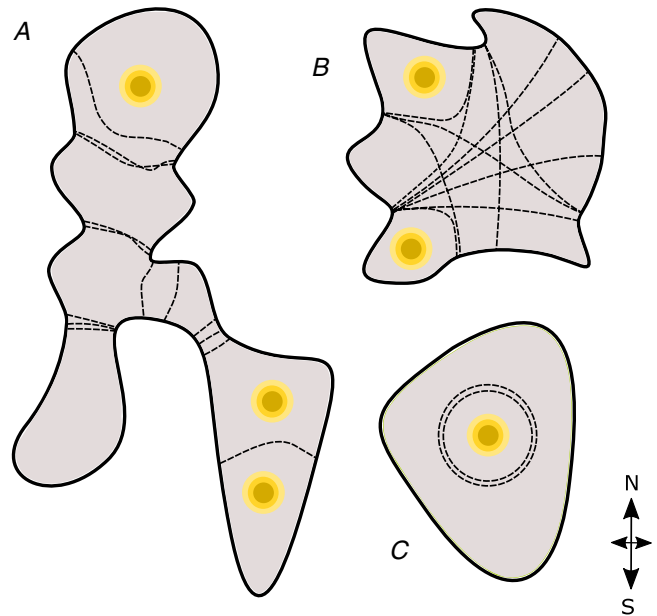

FIG. 19. Schematic diagram of isoglosses (dashed lines) for three language areas or "island nations." Yellow and ochre circles represent cities. Nation *A* supports dialect areas formed by coastal boundary shape and repulsion from cities. Nation *B* largely exhibits a continuum of variation apart from two somewhat indistinct city dialects. Nation *C* has a single city dialect, but without this city (or if the city were not sufficiently densely populated) it would have no linguistic variation due to its entirely convex boundary and evenly distributed population.

city, could not support more than one dialect, and would tend over time to lose isoglosses.

In some regions there are no dialect areas, only a continuum of variation [87], and in others clear dialects exist [4]. The above examples point to some general principles. In regions with low population density, a lack of major boundary indentations, and few large cities, we might expect isoglosses to position themselves in a less predictable way, creating language variation which would be perceived as a continuum by a traveler. The regular creation of new isoglosses (resetting the initial conditions of the model) through linguistic innovation or demographic instability could also disrupt the ordering processes. Narrow geographical regions, or where there are major boundary indentations, or dense population centers which push out linguistic change, are particularly susceptible to the formation of distinct dialects.

VII. TRANSITION REGIONS AND CURVATURE

We now derive analytical results that characterize the transition regions between variables and the effect of population density on the curvature of dialect domain boundaries.

To compute the gradients of transition regions, we consider a straight isogloss (with constant population density) in equilibrium between variants 1 and 2, given by the line $x = 0$. Because the isogloss is vertical the frequencies will not depend on y, so we write them $f_1(x)$ and $f_2(x)$ and note that $f_1(x) = 1 - f_2(x)$, so we need only consider the behavior of $f_1(x)$. For notational simplicity we define $f := f_1$, $m := m_1$ and $p(m) := p_1(\mathbf{m})$. The isogloss will form the midpoint of a transition region where the frequencies change smoothly between one and zero, and where $f_1'(0)$ measures the rate of this transition. In equilibrium, from Eq. (3) we have

$$\frac{\sigma^2}{2}\partial_x^2 p(m) = m - p(m).$$

Since $f = p(m)$, we have $(\sigma^2/2)f'' = p^{-1}(f) - f$. Note that f'' is shorthand for $\partial_x^2 f(x)$. If non-neutrality (conformity) is small, we may replace $p^{-1}(f)$ with its Taylor series about $\beta = 1$, neglecting $O[(\beta - 1)^2]$ terms:

$$\frac{\sigma^2}{2}f'' = (\beta - 1)f(1 - f)\ln\left(\frac{1 - f}{f}\right) + O[(\beta - 1)^2] \quad (34)$$

$$=: -\frac{dV}{df} + O[(\beta - 1)^2]. \quad (35)$$

Here, we define a "potential" function $V(f)$, allowing us to identify Eq. (34) as Newton's second law for the motion of a particle of mass $\sigma^2/2$ in a potential $V(f)$, where x plays the role of time, so that the total energy $E := (\sigma^2/4)f'(x)^2 + V(f(x))$ is independent of x [88].

Since V is defined by an indefinite integral, we can define $V(\frac{1}{2}) := 0$. As $x \to \pm\infty$, we require that $f'(x) \to 0$ and $f(x) \to 1$ or 0, so

$$E = \lim_{f \to 1}V(f) = \lim_{f \to 0}V(f) = \frac{4\ln 2 - 1}{24}(\beta - 1).$$

The magnitude of the frequency gradient at the origin, where $f(0) = \frac{1}{2}$, is, therefore,

$$|f'(0)| = \sqrt{\frac{4E}{\sigma^2}} \approx 0.545\frac{\sqrt{\beta - 1}}{\sigma}. \quad (36)$$

From this we see that weak non-neutrality and larger interaction range produce shallower gradients and therefore wider transition regions. As β approaches one, the transition region becomes increasingly wide and boundaries disintegrate, destroying the surface tension effect described in Fig. 3. Equation (36) is verified numerically for an "almost straight" isogloss in Fig. 20 (red dashed line).

We now relate the equilibrium shape of isoglosses to population density. Starting from our modified Allen-Cahn equation (9) for isogloss velocity, and introducing the local radius of curvature $R = 1/\kappa$, we see that when an isogloss is in equilibrium (having zero velocity)

$$R = -\frac{\rho}{2\nabla\rho \cdot \hat{\mathbf{g}}}, \quad (37)$$

where $\hat{\mathbf{g}}$ is a unit normal to the isogloss. We note a simple alternative derivation of this result based on the *dialect fraction* $F_{\mathcal{D}}(x, y)$ of a domain $\mathcal{D}$. We define this as the
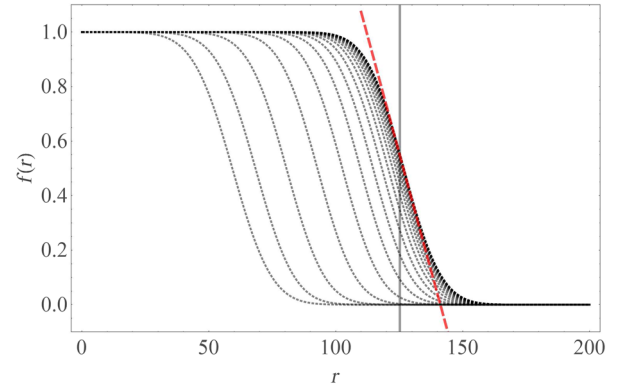

FIG. 20. Sequence of radial cross sections of the frequency of a linguistic variable whose initial domain is concentric with a city. Snapshots taken at times $t = 10, 30, 50, \ldots$ starting from initial radius $r = 55$. Model parameters $\sigma = 5$, $\beta = 1.1$. In this example, the city has radius $R_c = 50$ and peak density $\rho_1 = 3$. Vertical line gives theoretical stable radius $R_2 = 125.3$ computed from Eq. (41). Unstable radius [Eq. (40)] is $R_1 = 47.2$. Red dashed line gives the theoretical frequency gradient in transition region [Eq. (36)].

fraction of conversations that a speaker at a point (x, y) has with people whose home neighborhoods lie in $\mathcal{D}$:

$$F_{\mathcal{D}}(x, y) := \int_{\mathcal{D}} k(x, y; u, v) du dv. \qquad (38)$$

Let $P := (x^*, y^*)$ be a point on an isogloss with local radius of curvature $R \gg \sigma$, bounding some region $\mathcal{D}$. An intuitively appealing condition for language equilibrium is that the speaker at P should interact with equal numbers of speakers from within and without $\mathcal{D}$:

$$F_{\mathcal{D}}(x^*, y^*) = \frac{1}{2}. \qquad (39)$$

Using the saddle point method [47] to evaluate Eq. (38), we have

$$F_{\mathcal{D}} = \frac{1}{2} - \frac{\sigma}{\sqrt{2\pi}} \left(\frac{1}{2R} + \frac{\nabla \rho \cdot \hat{\mathbf{g}}}{\rho} \right) + O\left(\frac{\sigma^2}{R^2} \right).$$

Result Eq. (37) then follows from the equilibrium condition Eq. (39). From this we see that large gradients in population density can produce equilibrium isoglosses with higher curvature. To test this against our full evolution equation (3), consider a circular city with radius R_c having a Gaussian population distribution set in a unit uniform background:

$$\rho(x, y) = 1 + (\rho_1 - 1)e^{-(x^2+y^2)/2R_c^2}.$$

The constant $\rho_1 \geq 1$ gives the relative density of the city center compared to outlying areas. Consider a circular isogloss which is concentric with the city, then Eq. (37) has two solutions:

$$R_1 = R_c \sqrt{\frac{1}{2} - 2W_p\left(-\frac{e^{1/4}}{4(\rho_1 - 1)}\right)}, \qquad (40)$$

$$R_2 = R_c \sqrt{\frac{1}{2} - 2W_m\left(-\frac{e^{1/4}}{4(\rho_1 - 1)}\right)}, \qquad (41)$$

where W_p and W_m are two branches of Lambert's W function [89], $W(z)$, which solves $z = we^w$. The two solutions R_1 and R_2 are, respectively, decreasing and increasing functions of ρ_1. The stability of these solutions may be determined by noting that if $F_{\mathcal{D}} > \frac{1}{2}$, then $\mathcal{D}$ will expand, and contract if the inequality is reversed. From this we are able to determine that R_2 is stable, whereas R_1 is not. If a dialect domain begins with $R < R_1$, then it will shrink and disappear under the influence of surface tension, but if initially $R > R_1$, then the domain will expand or contract until $R = R_2$. This behavior is illustrated in Fig. 20, where

we see that our law Eq. (37) accurately predicts the stable radius produced by our evolution equation (3).

VIII. DISCUSSION AND CONCLUSION

A. Summary of results

Departing from the existing approaches of dialectology, we formulate a theory of how interactions between individual speakers control how dialect regions evolve. Much of what we demonstrate is a consequence of the similarity between dialect formation and the physical phenomenon of phase ordering. Humans tend to set down roots, and to conform to local speech patterns. These local patterns may be viewed as analogous to local crystal orderings in binary alloys [83] or magnetic domains [22]. As with phase ordering, surface tension is a dominant process controlling the evolution of dialect regions. However, important differences arise from the fact that human populations are not evenly distributed through space, and the geographical regions in which they live have irregular shapes. These two effects cause many different randomized early linguistic conditions to evolve toward a much smaller number of stable final states. For Great Britain we show that an ensemble of these final states can produce predicted dialect areas which are in remarkable agreement with the work of dialectologists.

Since language change is inherently stochastic, at small spatial scales we can only expect predictions of a statistical nature. At larger scales an element of deterministic predictability emerges. Within our model, all stochasticity arises from the randomization of initial conditions. The effect of this randomness is strongest in the early stages of language evolution, when the typical size of dialect domains is small. The superposition of isoglosses lacks a discernible pattern. This "tangle" of lines produces a continuum of language variation, with spatial correlations given by Séguy's curve. As surface tension takes hold, steered by variations in population density and system shape, isoglosses begin to bundle and the continuum resolves into distinct dialect regions. Both long-term population stability and large variations in population density play an important role. Without these ingredients isoglosses will not have time to evolve into smooth lines, or bundle.

The assumptions of our model are minimal, and clearly there are many additional complexities involved in language change which we have not captured; we discuss below how the model might be extended to account for some of these. Despite this simplicity, in addition to substantially reproducing Trudgill's predictions for English dialects [4], we are also able to explain the observation of both dialect continua and more sharply bounded dialect domains. We also provide an explanation for why boundary indentations (e.g., in coastline or in the border between different languages) are likely to collect

isoglosses [1]. We show that cities repel isoglosses, explaining the origin of the Rhenish fan [43], the wavelike spread of city language features [15], and the fact that many dialect patterns are centered on large urban areas. We explain why linguistic regions with high aspect ratio tend to develop striped dialect regions [77,78]. We compute an analytical form for Séguy's curve [2,5,6,39] which as yet has had no theoretical derivation. We also adapt this derivation to deal with a population center. We quantify how the width of a transition region [1] between dialects is related to the strength of conformity in individual language use and the typical geographical distances over which individuals interact. We show how to relate the curvature of stable isoglosses to population gradients. Finally, in Appendix B we show how incorporating an age distribution into the model can quantify the "'apparent time" [58] effect used by dialectologists to detect a linguistic change in progress. Given these findings, we suggest that the theoretical approach we present would be worth further investigation.

B. Missing pieces

The model we present is deliberately minimal: it allows us to see how much of what is observed can be explained by local interactions and conformity alone. This leads to a simple unified picture with surprising explanatory power. However, having chosen simplicity, we cannot then claim to provide a complete description of the processes at play. We now describe the missing pieces, indicating what effect we expect them to have, and how to include them.

1. Innovation

An important aspect of language change that is not explicitly encoded within our model is *innovation*: the creation of new forms of speech. We instead assume that there are a finite set of possible linguistic variables, and for each one, a finite set of alternatives, all of which are present in varying frequencies within the initial conditions of the model. Each alternative is equally attractive so that conformity alone decides its fate. A new variant cannot spontaneously emerge within the domain of another. The model therefore evolves toward increasing order and spatial correlation. Because of the presence of population centers this ordering process is eventually arrested creating distinct, stable domains. If innovation were allowed, then ordering would be interrupted by the initialization of new features, and Séguy's curve would reach a steady state where the rate of innovation (creating spatial disorder) balanced the rate of ordering.

For a local innovation to take hold, it must overcome local conformity, realized as surface tension and the "shrinking droplet" effect. Several mechanisms might allow this to happen: for example, young speakers must recreate their language using input from parents, peers, and other community members. This recreation process is inherently imperfect [19], and many interactions are between young speakers who are simultaneously assimilating their language. In this sense the young are weakly coupled to the current, adult speech community, and their language state is subject to fluctuations which may be sufficient to overcome local conformity for long enough to become established. As these speakers age their linguistic plasticity declines, older speakers die, and the change is cemented. Other mechanisms include speech modifications made to demonstrate membership of a social group, or a bias toward easier or more attractive language features. To understand mathematically the effect of innovation on spatial evolution, we might simply allow the creation of new variants, and then assign them a "fitness" relative to the existing population.

2. Interaction network

By selecting a Gaussian interaction kernel, and not distinguishing between different social groups, we are assuming that the social network through which language change is transmitted is only *locally* connected in a geographical sense but within each locality the social network is fully connected. That is, I will listen without prejudice to anyone regardless of age, sex, status, or ethnicity, as long as they live near my home. Research into 21st-century human mobility [90,91] and the work of linguists [55,58,61] indicates that both these assumptions are a simplification. Human mobility patterns, and by implication interaction kernels, exhibit heavy-tailed behavior (with an exponential cutoff at large distances). In our framework, an interaction kernel of this type, combined with densely populated cities would allow long-range connectivity between population centers. Long-range interactions in phase-ordering phenomena can have substantial effects on spatial correlations and domain sizes [92], and may effectively alter the spatial dimension of the system [93].

Further evidence for nonlocal networks is provided by the Frisian language, spoken in the Dutch province of Friesland. This has a "town Frisian" dialect [68], spoken in towns that are separated from each other by the Frisian countryside, where a different dialect is spoken: town Frisian is "distributed." Within the social network these towns are "local" (nearby) to each other. To incorporate this effect into our model, we must either reformulate our fundamental equation (3) to describe evolution on a more general network or generalize our interaction kernel to allow communication between cities. Further empirical evidence for nonlocal interactions is provided by hierarchical diffusion [94], where linguistic innovations spread between population centers, not necessarily passing through the countryside in between. Such a process motivated the creation of Trudgill's gravity model [14].

In addition, mobility data [91] show that individuals follow regular, repeating trajectories, introducing strong

heterogeneity within the set of individual interaction kernels. Social, as well as spatial, heterogeneity may also be important. For example, it has been shown theoretically [95] that the time required for two social groups to reach linguistic consensus is highly sensitive to the level of affinity that individuals have for their own group.

3. Linguistic space and dynamics

By using a set of *linguistic variables* we are treating dialects as points in vector space. Implicit in our dynamics are two assumptions. First, all transitions between variants are allowed, with probabilities given only by the frequencies with which the variables are used. Second, the evolution of different linguistic variables are mutually independent. There are cases where this is an incomplete description. A notable example is *chain shifting* in vowel sounds [58]. Linguists represents the set of possible vowel sounds as points in a two-dimensional domain where *closeness* of the tongue to the roof of the mouth and the position of the tongue's highest point (toward the *front* or *back* of mouth) form two orthogonal coordinate directions [96,97]. The vowel system of a language is then a set of points in this domain. If one vowel change leaves a gap in this system (an empty region of the domain), then other vowel sounds may shift to fill this gap, producing a chain of interconnected linguistic changes. Similarly, a change in one vowel to bring it closer to another can push it, and then others, out of their positions. A famous example is the "great vowel shift" [19] in England between the 14th and 17th centuries. Another example concerns changes that spread to progressively more general linguistic (as opposed to geographical) contexts [81]. If we have several variables, each signifying the presence of the change in a different context, then it is clear that the frequency of one variable can influence the frequency of another, contradicting our assumption that variables are independent.

The fact that linguistic variables are sometimes dependent upon one another means that, within our model, p_i, which relates the past use of some variable to the current frequency of its ith variant, via the relationship $f_i(x, y, t) := p_i[\mathbf{m}(x, y, t)]$, should sometimes depend on the use of other variables, and might be adapted to capture more than just conformity.

C. Conclusion

We conclude by noting that a major theme of the book *War and Peace* by Tolstoy is the idea that history is determined not by great individuals but rather by millions of small choices made by the people.

To elicit the laws of history we must leave aside kings, ministers, and generals, and select for study the homogeneous, infinitesimal elements which influence the masses [98].

As pointed out by Vitány [99], Tolstoy was, in modern terms, advocating the formulation of a statistical mechanics of history. The work we present is an attempt to formulate such a theory for the spatial history of language. Because of its simplicity, dealing only with copying and movement, our model may apply more broadly to other forms of culture.

APPENDIX A: DISCRETIZED EVOLUTION EQUATION

Here, we present the computational scheme used for solving our evolution equation (3). This is aimed at researchers having some familiarity with computer programing, such as linguists interested in quantitative approaches. It can also be implemented using only a spreadsheet (see Supplemental Material [38]). The discretized version of evolution equation (3) also provides a greater intuitive understanding of its continuous counterpart. For simplicity, we consider the $V = 2$ model and define $f := f_1$, $m := m_1$, $p(m) := p_1(\mathbf{m})$, and note that we need only consider the evolution of m and f because $f_2 = 1 - f_1 = 1 - f$.

We begin by rewriting our evolution equation (3) in terms of the memory and frequency fields

$$\partial_t m = f - m + \frac{\sigma^2}{2\rho} \nabla^2(\rho f), \qquad (A1)$$

where

$$f = p(m) = \frac{m^\beta}{m^\beta + (1 - m)^\beta}. \qquad (A2)$$

To solve Eq. (A1) on a computer, we discretize space into a rectangular grid of square sites. We let the side of each grid square define one unit of length. The interaction range used in the computer calculation should be expressed in these units. That is, if the side of a grid square is a-km long, and the real-world interaction range is σ km, then the interaction range used in the computer should be $\sigma_c := \sigma/a$. We choose a so that $\sigma_c > 1$, so speakers interact over distances greater than one square. The subscript c distinguishes the interaction range measured in computer grid units from the

interaction range in km. For each site we store three quantities, ρ_{ij}, f_{ij}, m_{ij}, where the subscripts i, j are horizontal and vertical indices related to spatial position by $x = ia$ and $y = ja$. These quantities are our approximations to the values of the fields ρ, f, m at the centers of sites, and are stored in three arrays. Intuitively, we think of each site as containing a group of $a^2 \rho_{ij}$ speakers, each of whom uses variant 1 with frequency f_{ij}. The linguistic domain of interest is the set of sites with nonzero population density. Within the domain, sites with one or more nearest neighbors with zero population are referred to as *boundary* sites, otherwise they are *bulk* sites. Sites which are not part of the domain are never updated, and it is useful to include a border of such sites around the edge of the rectangular grid.

The scheme we present is based on approximating the Laplacian $\nabla^2 = \partial_x^2 + \partial_y^2$ using a central finite difference approximation for the derivatives ∂_x^2 and ∂_y^2. Let g be an arbitrary function defined at every site. We define a *local average* at each grid point:

$$\langle g \rangle_{ij} = \frac{1}{4}(g_{i+1,j} + g_{i-1,j} + g_{i,j+1} + g_{i,j-1}). \quad (A3)$$

This is just the average of the values of g at the four nearest neighbors of (i, j). The Laplacian is then approximated:

$$\nabla^2 g_{ij} \approx 4(\langle g \rangle_{ij} - g_{ij}). \quad (A4)$$

This follows from the finite difference approximation $\partial_x^2 g \approx g_{i+1,j} - 2g_{ij} + g_{i-1,j}$, the effectiveness of which depends on g varying slowly between sites. From Eq. (A4) we see that $\nabla^2 g$ measures the extent to which g differs from the average of its neighbors. If $\nabla^2 g < 0$, then g exceeds the local average, and is less than the local average when the inequality is reversed.

We now introduce a small discrete time step δt and write $\Delta m_{ij} := m_{ij}(t + \delta t) - m_{ij}(t)$ for the change in the memory field over the time interval $[t, t + \delta t]$. We note also that provided the grid is sufficiently fine, then at bulk sites $\rho_{ij} \approx \langle \rho \rangle_{ij}$, so, making use of Eqs. (A1) and (A4), we have

$$\Delta m_{ij} \approx \left[(f_{ij} - m_{ij}) + 2\sigma^2 \left(\frac{\langle \rho f \rangle_{ij}}{\langle \rho \rangle_{ij}} - f_{ij} \right) \right] \delta t. \quad (A5)$$

At each time step Eq. (A5) is used to update the stored values of m_{ij} for all sites in the linguistic domain, after which the frequencies can also be updated using $f_{ij} = p(m_{ij})$. The quantity $\langle \rho f \rangle_{ij} / \langle \rho \rangle_{ij}$ is the average of frequencies at neighboring sites, weighted in proportion to their populations. Using $\langle \rho \rangle_{ij}$ rather than ρ_{ij} in the denominator ensures that these weights sum to one. This serves two purposes. First, it avoids the need for an additional condition at boundary sites (intuitively, speakers in

boundary sites simply shift attention from empty neighboring squares to those which are part of the linguistic domain, consistent with the original definition of the weighted interaction kernel). Second, it ensures that spatially constant memory and frequency fields constitute a fixed point of the dynamics. Rule Eq. (A5) is an explicit scheme and as such its stability requires that δt be chosen sufficiently small. In the case of zero conformity ($\beta = 1$) and constant population density, the von Neumann stability condition [63] is $\delta t < 1/\sigma^2$. This serves as a guide to find δt sufficiently small for our scheme to converge. For the density fields and conformity values we use in this paper, we find that $\delta t < 1/(4\sigma_c^2)$ is more than sufficient.

We conclude this section by explaining the linguistic meaning of the terms on the right-hand side of Eq. (A5). The first term, $f_{ij} - m_{ij}$, drives conformity. If $m_{ij} > \frac{1}{2}$, this term is positive, driving the memory further towards $m_{ij} = 1$ where all speakers use variant 1. Otherwise, if $m_{ij} < \frac{1}{2}$, the memory is driven towards zero where no speakers do. The second term, $\langle \rho f \rangle_{ij} / \langle \rho \rangle_{ij} - f_{ij}$, acts to equalize speech use in the local area. If variant 1 is used more at (i, j) than in the surrounding squares, then this term acts to reduce its use in (i, j). If variant 1 is used relatively less at (i, j), the term has the opposite effect.

APPENDIX B: INCORPORATING AGE INTO THE MODEL

In order to experimentally detect a linguistic change in progress, ideally one would like to survey the same population of individuals, or a representative sample of the population, at two or more points in time [1]. Such longitudinal studies may be practically difficult to carry out, so linguists have made use of the assumption that speech patterns are acquired mainly in the early part of people's lives. The speech of a 50-year old today should therefore reflect the speech of a 30-year old 20 years ago. It should be noted though that speech patterns can change throughout life [44], although more slowly in older speakers. A linguistic change detected by observing different speech patterns in the young and old is said to have been observed in *apparent time* [58,59]. A famous example of apparent time is the replacement of the term *chesterfield* (meaning an upholstered multiple-person seat) in Canadian speech with the fashionable American term *couch* [100]. In this case the use of *couch* was shown to decrease sigmoidally from ≈85% amongst teenagers to ≈5% among those in their eighties. The apparent time theory has been tested by comparing language surveys taken at different times and comparing predictions based on apparent time in the earlier sample with the observations made in the later one [101]. We note that differences between speech patterns between the generations do not always indicate a linguistic change in progress [44]. For example, the use of some speech forms may change systematically with age in the same way,

generation after generation, so that the community as a whole is in a stable state [55].

We now give a simple illustration of how age and apparent time can be incorporated in our model. For simplicity, we consider the progress of a straight isogloss between two variants, driven by a slowly declining population density. This density variation is equivalent to a social bias toward one variable which we call the *new variant*. We let $f(x, t)$ be the frequency with which the new (spreading) variable is used at position x and time t. Note that there is no y dependence due to symmetry. To distinguish between young and old we introduce an age density distribution $\alpha(a)$ giving the fraction of individuals within the age bracket $[a_1, a_2]$ as

$$\int_{a_1}^{a_2} \alpha(a) da. \tag{B1}$$

Using this distribution we modify our original model to account for the fact that individuals have been exposed only to the linguistic information available in their lifetime. The memory of a speaker with age a is therefore defined as

$$\mu_a(x, t) := \int_0^a \frac{e^{-s/\tau}}{\tau(1 - e^{-a/\tau})} \left[\int_{\mathbb{R}} k(x; u) f(u, t - s) du \right] ds. \tag{B2}$$

Note that as $a \to \infty$ this definition coincides with our original definition Eq. (2) of memory. In the interests of analytical tractability, we consider the limit of small memory decay rate ($\tau \to \infty$) in which case linguistic memory is a simple "bus stop" average over life history:

$$\mu_a(x, t) = \frac{1}{a} \int_0^a \left[\int_{\mathbb{R}} k(x; u) f(x, t - s) du \right] ds. \tag{B3}$$

We also take the limit of total conformity $\beta \to \infty$ so that language is chosen according to a simple majority rule. We consider an exponentially decaying population density

$$\rho(x) = e^{-\epsilon x}, \tag{B4}$$

where $\epsilon \ll 1$. The weighted interaction kernel for this density is then

$$k(x; u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(u - x + \epsilon\sigma^2)^2}{2\sigma^2} \right\} \tag{B5}$$

$$=: k(u - x). \tag{B6}$$

Notice that the effect of the decaying population density is to shift the interaction kernel to the left so that more attention is paid to language use on that side of the listener. To compute the isogloss velocity, we define

$$\eta(x) := \begin{cases} 1 & \text{if } x < 0 \\ \frac{1}{2} & \text{if } x = 0 \\ 0 & \text{if } x > 0 \end{cases} \tag{B7}$$

and prepare the system with the initial condition

$$f(x, t < t_0) = \eta(x - \Lambda), \tag{B8}$$

where Λ is the initial location of the isogloss. Because each speaker listens more to the speakers on their left, the isogloss will travel right. In the limit $\beta \to \infty$ then when the memory $\mu_a(x, t)$ of a speaker, with $x > \Lambda$, reaches $\frac{1}{2}$, they will switch linguistic variables. The motion of the isogloss will then take the form of a traveling wave formed from a superposition of traveling step functions, one for each age,

$$f(x, t) = \int_0^\infty \alpha(a) \eta[x - vt + \Lambda(a)] da, \tag{B9}$$

with the function $\Lambda(a) > 0$ and the velocity v to be determined. According to Eq. (B9), at $t_a := \Lambda(a)/v$, a speaker at the origin with memory of length a will be on the point of a switching variable, so that

$$\mu_a(0, t_a) = \frac{1}{2} = \int_0^a \frac{ds}{a} \int_0^\infty da' \int_{\mathbb{R}} dy \alpha(a') k(y)$$
$$\times \eta[y + vs - \Lambda(a) + \Lambda(a')]$$
$$= \int_0^a \frac{ds}{a} \int_0^\infty da' \alpha(a') \int_{-\infty}^{\Lambda(a) - \Lambda(a') - vs} k(y) dy$$
$$=: \int_0^a \frac{ds}{a} \int_0^\infty \alpha(a') K[\Lambda(a) - \Lambda(a') - vs] da'. \tag{B10}$$

Here, we introduce the cumulative K of the interaction kernel:

$$K(z) := \int_{-\infty}^z k(y) dy. \tag{B11}$$

As $\epsilon \to 0$, the isogloss velocity must also tend to zero. The quantity

$$\Delta(a_1, a_2) := \Lambda(a_1) - \Lambda(a_2) \tag{B12}$$

gives the distance between the step functions for speakers with ages a_1 and a_2 as $|\Delta(a_1, a_2)|$, and this separation must also tend to zero as $\epsilon \to 0$. We can therefore compute a series expansion for v in powers of ϵ by expanding the cumulative interaction kernel $K(z)$ in Eq. (B10) about $z = 0$ and $\epsilon = 0$. To lowest order we have

$$K(z) = \frac{1}{2} + \frac{z}{\sqrt{2\pi}\sigma} + \frac{\epsilon\sigma}{\sqrt{2\pi}} + o(z) + o(\epsilon). \tag{B13}$$

Substituting this approximation into Eq. (B10), we have

$$\int_0^\infty \frac{\Delta(a, a')}{\sigma} \alpha(a') da' - \frac{av}{2\sigma} + \epsilon\sigma = 0. \qquad \text{(B14)}$$

It is straightforward to verify that this equation has the solution

$$v = \frac{2\sigma^2\epsilon}{\bar{a}}, \qquad \text{(B15)}$$

$$\Delta(a, a') = \frac{(a - a')\sigma^2\epsilon}{\bar{a}}, \qquad \text{(B16)}$$

where $\bar{a}$ is the mean age of the population:

$$\bar{a} := \int_0^\infty a\alpha(a) da. \qquad \text{(B17)}$$

If the oldest speaker has age A, then the width of the transition region is $\Delta(A, 0) = A\sigma^2\epsilon/\bar{a}$. We provide a concrete example using a population "pyramid" age distribution, cut off exponentially at low ages to account for the fact that very young speakers listen to, but do not influence, others. Letting a_0 be the low age cutoff, we define

$$\alpha(a) = \frac{1}{\mathcal{C}}(1 - e^{-a/a_0})(A - a), \qquad \text{(B18)}$$

where $\mathcal{C}$ is a normalizing constant. An example of the traveling wave Eq. (B9) generated by this age distribution is illustrated in Fig. 21. Also shown are the results of a numerical implementation of the full model with a discretized version of the age distribution Eq. (B18). This discretization is necessary in order to implement the model numerically, because the memory of each age of speaker must be individually stored.

Finally, we consider the likely outcome of experimentally sampling the use of language within this model. We let x_0 be the left boundary of the transition region (the oldest speaker at x_0 is just about to switch variable). We then define the indicator function of the event that a speaker of age a, located at position x, is using the new variant:

$$q_x(a) := \begin{cases} 1 & \text{if } x - x_0 < \Delta(A, a) \\ 0 & \text{otherwise.} \end{cases} \qquad \text{(B19)}$$

Consider a sample of speakers with home locations X, normally distributed around some average position $x_0 + h$: $X \sim \mathcal{N}(x_0 + h, \omega^2)$. The probability that a speaker of age a within this sample will use the new variant is then the expectation of $q_X(a)$ over the position X:
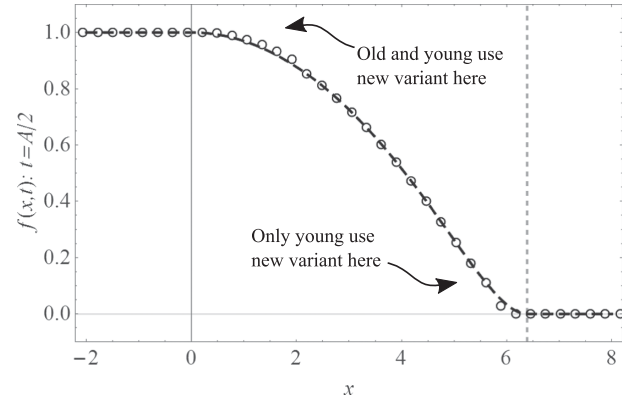

FIG. 21. Thick dashed line shows theoretical frequency of new variant for age distribution Eq. (B18)—using linear approximations Eq. (B15) and (B16) for $v(a)$ and $\Lambda(a)$, and assuming $\Lambda(0) = 0$—when $a_0 = 10$, $A = 90$, $\sigma = 5$, $\epsilon = 0.1$ at time $t = A/2 = 45$ (chosen so that oldest speaker at $x = 0$ has just switched variable). For these parameter values $\bar{a} = 35.23$. Vertical dashed line is drawn at location of youngest adopter of new variable (giving width of transition region). Open circles give frequency values using an age distribution discretized into two-year bins, computed by numerically evolving the full model.

$$\mathbf{E}[q_X(a)] = \frac{1}{2}\text{erfc}\left(\frac{h - \Delta(A, a)}{\sqrt{2}\omega}\right) =: \bar{q}(a; h, \omega). \qquad \text{(B20)}$$

An example of this distribution is illustrated in Fig. 22. In this example, the mean sample location is the center of the transition region, and we see that uptake of the new variant exhibits the characteristic "S-shaped" age distribution seen in studies of linguistic change observed in apparent time [1,58,100].

We conclude by noting that this extension of the model to include different memory lengths should be seen as a toy model of the effect of age on language change. The fact that
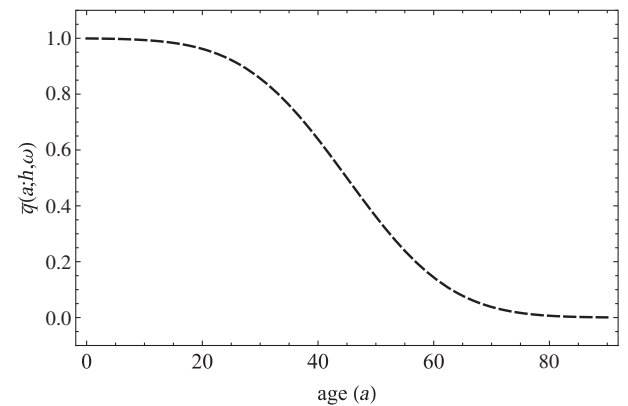

FIG. 22. Expected frequency Eq. (B20) with which new variant is used by speakers of different ages from a random sample. Mean and variance of speaker locations relative to left boundary x_0 of transition region are $h = \frac{1}{2}\Delta(0, A)$ and $\omega^2 = 1$. Model parameter values are $a_0 = 10$, $A = 90$, $\sigma = 5$, $\epsilon = 0.1$. For these parameter values, $\bar{a} = 35.23$.

older speakers tend to take longer to change their language use is captured purely by the length of their memories. In reality, the influence on a given speaker of the language they are exposed to at different stages of life will be much more complicated than our simple model [102]. For example, if language use were determined entirely during early life, then the *forgetting curve* should be peaked during these early years—in effect, the linguistic memory should stop "recording" once a speaker's youth has ended. Each speaker will respond differently to what they hear, so the forgetting curve will not be identical for every speaker. Such heterogeneity amongst speakers is straightforward to incorporate, but at the cost of tractability. The advantage of the simple approach is to illustrate the apparent time effect in an analytically simple way.

APPENDIX C: ALLEN-CAHN EQUATION AND OHTA-JASNOW-KAWASAKI THEORY

Here, we derive an analogue of the Allen-Cahn equation [83] for the velocity of an isogloss, before deriving a modified Ohta-Jasnow-Kawasaki equation [84], which provides a simplified method for understanding the evolution of spatial correlations in the model. In the binary variant case ($V = 2$), we have that $m_1 = 1 - m_2$, so after defining $p(m_1) := p_1(\mathbf{m})$ and $f := f_1$, then $f = p(m_1)$. In terms of f, our evolution equation (3) may be written as

$$\partial_t f = p'[p^{-1}(f)]\left\{ f - p^{-1}(f) + \frac{\sigma^2}{2\rho}\nabla^2(\rho f)\right\}. \quad (C1)$$

Following Refs. [20,83], we introduce a unit vector $\hat{\mathbf{g}}$, normal to the isogloss at a point P. We let g be the displacement from P in the direction of $\hat{\mathbf{g}}$. At the isogloss we have

$$\nabla f = \left(\frac{\partial f}{\partial g}\right)_t \hat{\mathbf{g}},$$

$$\nabla^2 f = \left(\frac{\partial^2 f}{\partial g^2}\right)_t + \left(\frac{\partial f}{\partial g}\right)_t \nabla \cdot \hat{\mathbf{g}}.$$

We also need the cyclic relation:

$$\left(\frac{\partial t}{\partial f}\right)_g \left(\frac{\partial f}{\partial g}\right)_t \left(\frac{\partial g}{\partial t}\right)_f = -1. \quad (C2)$$

The Laplacian term in Eq. (C1) may be expanded as follows:

$$\frac{\nabla^2(\rho f)}{\rho} = \nabla^2 f + f\frac{\nabla^2\rho}{\rho} + 2\frac{\nabla\rho \cdot \nabla f}{\rho} \quad (C3)$$

$$= \left(\frac{\partial^2 f}{\partial g^2}\right)_t + \left(\frac{\partial f}{\partial g}\right)_t \nabla \cdot \hat{\mathbf{g}} + f\frac{\nabla^2\rho}{\rho} + 2\frac{\nabla\rho \cdot \nabla f}{\rho}. \quad (C4)$$

For constant density, the equilibrium configuration of the isogloss is a straight line, so the curvature $\kappa := \nabla \cdot \hat{\mathbf{g}}$ is zero, and f depends only on the displacement g. Making use of Eqs. (C1) and (C4), we see that the equilibrium equation for f is in this case

$$\frac{\sigma^2}{2}\frac{d^2 f}{dg^2} = p^{-1}(f) - f. \quad (C5)$$

We now make the assumption that out of equilibrium, if the curvature is low, and the density slowly varying, then the profile of the transition region around the isogloss takes its equilibrium form. We also recall our assumption in our derivation of the full evolution equation (3) that $|\nabla^2\rho|/\rho \ll \sigma^2$. The evolution equation (C1) may then be written as

$$\partial_t f = \frac{\sigma^2}{2} p'[p^{-1}(f)]\left\{ \left(\frac{\partial f}{\partial g}\right)_t \nabla \cdot \hat{\mathbf{g}} + 2\frac{\nabla\rho \cdot \nabla f}{\rho}\right\} \quad (C6)$$

$$= \sigma^2 p'[p^{-1}(f)]\left(\frac{\partial f}{\partial g}\right)_t\left\{\frac{\nabla \cdot \hat{\mathbf{g}}}{2} + \frac{\nabla\rho \cdot \hat{\mathbf{g}}}{\rho}\right\}. \quad (C7)$$

Making use of relation Eq. (C2), we have

$$\left(\frac{\partial g}{\partial t}\right)_f = -\sigma^2 p'[p^{-1}(f)]\left\{\frac{\kappa}{2} + \frac{\nabla\rho \cdot \hat{\mathbf{g}}}{\rho}\right\}. \quad (C8)$$

Since $(\partial g/\partial t)_f$ is the isogloss velocity, and at the isogloss we have $f = \frac{1}{2}$, then

$$v = -\sigma^2\beta\left\{\frac{\kappa}{2} + \frac{\nabla\rho \cdot \hat{\mathbf{g}}}{\rho}\right\}. \quad (C9)$$

To obtain spatial correlation functions between different modal linguistic variables, we apply the Ohta-Jasnow-Kawasaki method [84]. We adapt the description of OJK's analysis given in Bray [20] to include population density effects. As we describe in Sec. V, we label the two alternatives for a particular variable as -1 and 1 and introduce a smoothly varying auxiliary field $m(x, y, t)$ which gives the modal (most common) variant ϕ_i of variable i as $\phi_i(m) = \text{sgn}(m)$. The unit vector $\hat{\mathbf{g}}$ may then be written as

$$\hat{\mathbf{g}} = \frac{\nabla m}{|\nabla m|}, \quad (C10)$$

from which we see that the isogloss velocity is

$$v = \sigma^2\beta\left(\frac{-\nabla^2 m + \sum_{i,j}\hat{g}_i\hat{g}_j\partial_i\partial_j m - 2(\nabla\rho.\nabla m)/\rho}{2|\nabla m|}\right), \quad (C11)$$

where $i, j \in \{x, y\}$. In a reference frame attached to the interface,

$$\frac{dm}{dt} = 0 = \frac{\partial m}{\partial t} + \mathbf{v} \cdot \nabla m. \quad \text{(C12)}$$

Since $\mathbf{v} \| \nabla m$, then $\mathbf{v} \cdot \nabla m = v|\nabla m|$ and $\partial_t m = -v|\nabla m|$, so

$$\frac{\partial m}{\partial t} = \frac{\sigma^2 \beta}{2}\left(\nabla^2 m - \sum_{i,j} \hat{g}_i \hat{g}_j \partial_i \partial_j m + 2\frac{\nabla\rho.\nabla m}{\rho}\right). \quad \text{(C13)}$$

This is the OJK equation, modified to include variable population density. As OJK did, we now assume that the direction $\hat{\mathbf{g}}$ is uniformly distributed over the system, and we replace $\hat{g}_i \hat{g}_j$ with its circular mean $\frac{1}{2}\delta_{ij}$, giving

$$\frac{\partial m}{\partial t} = \sigma^2 \beta \left(\frac{\nabla^2 m}{4} + \frac{\nabla\rho.\nabla m}{\rho}\right). \quad \text{(C14)}$$

This is our modified Ohta-Jasnow-Kawasaki equation.

APPENDIX D: COMPARISON WITH A SIMULATION OF NERBONNE

We note a link between curve Eq. (19) and a simple model simulated by Nerbonne [39], but not characterized analytically. The model consists of a line of discrete spatial points, with a single reference site at one end representing a city. In contrast to our model, all sites initially use an identical set of N ($= 100$ in Ref. [39]) binary linguistic variables. Evolution of language use is simulated at each site by repeatedly selecting a variable at random and then changing the state of the variable with probability $\frac{1}{2}$. For a site at distance r from the city, the number of repeats of this randomization process is defined as $n(r) := \lfloor Cr^\alpha \rfloor$, where the constant α measures the spatial decline of the "influence" of the city with r. Larger values of $n(r)$ imply a greater level of noisy evolution and therefore a lower influence of the city. Linguistic distance in this model, after each site has received its $n(r)$ updates, is given by

$$l(r) = \frac{1}{2}[1 - (1 - N^{-1})^{n(r)}] \approx \frac{1}{2}[1 - e^{-n(r)/N}]. \quad \text{(D1)}$$

In Ref. [39], two values of α are tested, $\alpha = 1, 2$, and the quadratic case is identified as being consistent with Trudgill's macroscopic gravity model. However, we emphasize that the two models do not make predictions that can be directly compared: in the microscopic model [39] no indication is given of how two centers of influence would compete. Of interest is the fact that the $\alpha = 2$ case coincides with our prediction for large r. However, this value of α is rejected in Ref. [39] on the basis of its sigmoidal shape for small r. Because of the presence of the

inverse sine function, this behavior is not present in our version of Séguy's law [Eq. (19)]. We note also that the dynamics of Ref. [39] reduce order in the system, whereas our model leads to increasingly ordered states.

[1] J. K. Chambers and P. Trudgill, *Dialectology* (Cambridge University Press, Cambridge, England, 1998).

[2] W. Heeringa and J. Nerbonne, *Dialect Areas and Dialect Continua*, Lang. Var. Change **13**, 375 (2001).

[3] L. Bloomfield, *Language* (Holt, Rinehart and Winston, New York, 1933).

[4] P. Trudgill, *The Dialects of England* (Blackwell, Hoboken, NJ, 1999).

[5] J. Séguy, *La Relation Entre la Distance Spatiale et la Distance Lexicale*, Rev. Ling. Romane **35**, 335 (1971).

[6] J. Séguy, *La Dialectomtrie Dans l'Atlas Linguistique de la Gascogne*, Rev. Ling. Romane **37**, 1 (1973).

[7] M. Wieling and J. Nerbonne, *Advances in Dialectometry*, Annu. Rev. Ling. **1**, 243 (2015).

[8] B. Kessler, in *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 1995), pp. 60–67.

[9] V. Levenshtein, *Binary Codes Capable of Correcting Deletions, Insertions and Reversals*, Dokl. Akad. Nauk SSSR **163**, 845 (1965).

[10] R. G. Shackelton, Jr., *Phonetic Variation in the Traditional English Dialects. A Computational Analysis*, J. Engl. Ling. **35**, 30 (2007).

[11] S. Embleton, D. Uritescu, and E. S. Wheeler, *Defining Dialect Regions with Interpretations: Advancing the Multidimensional Scaling Approach*, Literary Ling. Comput. **28**, 13 (2013).

[12] M. Wieling and J. Nerbonne, *Bipartite Spectral Graph Partitioning for Clustering Dialect Varieties and Detecting Their Linguistic Features*, Comput. Speech Lang. **25**, 700 (2011).

[13] M. Wieling and J. Nerbonne, GABMAP: *A Web Application for Dialectology*, Dialectologia **II**, 65 (2011).

[14] P. Trudgill, *Linguistic Change and Diffusion: Description and Explanation in Sociolinguistic Dialect Geography*, Lang. Soc. **3**, 215 (1974).

[15] W. Wolfram and N. Schilling-Estes, *Dialectology and Linguistic Diffusion*, in *The Handbook of Historical Linguistics*, edited by B. D. Joseph and R. D. Janda (Blackwell Publishing, Malden, 2003), pp. 713–725.

[16] W. Labov, *Social Dialectology in Honour of Peter Trudgill* (John Benjamins Publishing Company, Amsterdam, 2003), pp. 9–22.

[17] G. Bailey, T. Wikle, J. Tillery, and L. Sand, *Some Patterns of Linguistic Diffusion*, Lang. Var. Change **5**, 359 (1993).

[18] J. N. Stanford and L. A. Kenney, *Revisiting Transmission and Diffusion: An Agent-Based Model of Vowel Chain Shifts Across Large Communities*, Lang. Var. Change **25**, 119 (2013).

[19] D. Crystal, *The Cambridge Encyclopedia of the English Language* (Cambridge University Press, Cambridge, England, 2003).

[20] A. J. Bray, *Theory of Phase-Ordering Kinetics*, Adv. Phys. **43**, 357 (1994).

[21] K. Barros, P. L. Krapivsky, and S. Redner, *Freezing into Stripe States in Two-Dimensional Ferromagnets and Crossing Probabilities in Critical Percolation*, Phys. Rev. E **80**, 040101 (2009).

[22] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, England, 2010).

[23] J. Burridge and S. Kenney, *Birdsong Dialect Patterns Explained Using Magnetic Domains*, Phys. Rev. E **93**, 062402 (2016).

[24] M. Gerlach and E. G. Altmann, *Stochastic Model for the Vocabulary Growth in Natural Languages*, Phys. Rev. X **3**, 021006 (2013).

[25] A. M. Petersen, J. N. Tenenbaum, S. Havlin, H. E. Stanley, and M. Perc, *Languages Cool as They Expand: Allometric Scaling and the Decreasing Need for New Words*, Sci. Rep. **2**, 943 (2012).

[26] J. C. Willis and G. U. Yule, *Some Statistics of Evolution and Geographical Distribution in Plants and Animals, and Their Signature*, Nature (London) **109**, 177 (1922).

[27] M. E. J. Newman, *Power Laws, Pareto Distributions and Zipf's Law*, Contemp. Phys. **46**, 323 (2005).

[28] M. Gerlach and E. G. Altmann, *Scaling Laws and Fluctuations in the Statistics of Word Frequencies*, New J. Phys. **16**, 113010 (2014).

[29] M. Newman, *Networks: An Introduction*, (Oxford, New York, 2010).

[30] J. P. Gleeson, K. P. O'Sullivan, R. A. Banos, and Y. Moreno, *Effects of Network Structure, Competition and Memory Time on Social Spreading Phenomena.*, Phys. Rev. X **6**, 021019 (2016).

[31] R. Albert, H. Jeong, and A. L. Barabasi, *Error and Attack Tolerance of Complex Networks*, Nature (London) **406**, 378 (2000).

[32] M. Perc, *Evolution of the Most Common English Words and Phrases Over the Centuries*, J. R. Soc. Interface **9**, 3323 (2012).

[33] V. V. Loreto, A. Baronchelli, A. Mukherjee, A. Puglisi, and F. Tria, *Statistical Physics of Language Dynamics*, J. Stat. Mech. (2011) P04006.

[34] G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane, *Utterance Selection Model of Language Change*, Phys. Rev. E **73**, 046118 (2006).

[35] M. R. D'Orsogna and M. Perc, *Statistical Physics of Crime: A Review*, Phys. Life Rev. **12**, 1 (2015).

[36] Z. Wanga, C. T. Bauch, Samit Bhattacharyya, A. d'Onofrio, P. Manfredi, M. Perc, N. Perra, M. Salathé, and D. Zhaol, *Statistical Physics of Vaccination*, Phys. Rep. **664**, 1 (2016).

[37] C. Castellano, S. Fortunato, and V. Loreto, *Statistical Physics of Social Dynamics*, Rev. Mod. Phys. **81**, 591 (2009).

[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.7.031008 for animated versions of Fig. 6, Fig. 10, Fig. 18, and a spreadsheet implementation of the discretized evolution equation derived in Appendix A.

[39] J. Nerbonne, *Measuring the Diffusion of Linguistic Change*, Phil. Trans. R. Soc. B **365**, 3821 (2010).

[40] M. R. Mehl, S. Vazire, N. Ramirez-Esparza, R. B. Slatcher, and J. W. Pennebaker, *Are Women Really More Talkative Than Men?*, Science **317**, 82 (2007).

[41] J. K. Chambers, *Dialect Acquisition*, Language **68**, 673 (1992).

[42] J. Siegel, *Second Dialect Acquisition* (Cambridge University Press, Cambridge, England, 2010).

[43] P. Bloom, *How Children Learn the Meanings of Words* (MIT Press, Cambridge, MA, 2000).

[44] G. Sankoff and H. Blondeau, *Language Change Across the Lifespan: /r/ in Montreal French*, Language **83**, 560 (2007).

[45] L. Averell and A Heathcote, *The Form of the Forgetting Curve and the Fate of Memories*, J. Math. Psychol. **55**. 25 (2011).

[46] R. A. Blythe, *Neutral Evolution: A Null Model for Language Dynamics*, Adv. Complex Syst. **15**, 1150015 (2012).

[47] J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Addison-Wesley, Reading, MA, 1971).

[48] W. Croft, *Explaining Language Change: An Evolutionary Approach* (Longman, Harlow, England, 2000).

[49] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, England, 1983).

[50] P. A. P Moran, *Random Processes in Genetics*, Proc. Cambridge Philos. Soc. **54**, 60 (1958).

[51] S. Asch, *Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority*, Psychol. Monogr. **70**, 1 (1956).

[52] R. Bond and P. B. Smith, *Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task*, Psychol. Bull. **119**, 111 (1996).

[53] T. J. H. Morgan, L. E. Rendell, M. Ehn, W. Hoppitt, and K. N. Laland, *The Evolutionary Basis of Human Social Learning*, Proc. R. Soc. B **279**, 653 (2012).

[54] D. A. Nelson, *Song Overproduction, Selective Attrition and Song Dialects in the White-Crowned Sparrow*, Animal Behaviour **60**, 887 (2000).

[55] P. Trudgill, *Sociolinguistics* (Penguin, London, 2000).

[56] W. Labov, *The Social Stratification of English in New York City* (Cambridge University Press, Cambridge, England, 2006).

[57] A. S. Kroch, *Toward a Theory of Social Dialect Variation*, Lang. Soc. **7**, 17 (1978).

[58] W. Labov, *Principles of Linguistic Change* (Blackwell, Malden, MA, 2001).

[59] W. Labov, *The Social Motivation of a Sound Change*, Word and Information Processing **19**, 273 (1963).

[60] W. Wolfram and N. Schilling-Estes, in *Sociolinguistic Variation: Data, Theory, and Analysis: Selected Papers from NWAV 23 at Stanford* (CSLI, Stanford, CA, 1996), pp. 69–82.

[61] W. Labov, *Sociolinguistic Patterns* (University of Pennsylvania Press, Philadelphia, PA, 1972).

[62] D. MacAulay, *The Celtic Languages* (Cambridge University Press, Cambridge, England, 2008).

[63] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes* (Cambridge University Press, Cambridge, England, 2007).

[64] Office for National Statistics: 2011 Census aggregate data, UK Data Service, 2016, http://dx.doi.org/10.5257/census/aggregate-2011-1.

[65] M. Tranter, D. Robineau, and G. Goodman, National Travel Survey: 2014 Report (Department of Transport, United Kingdom, 2014), https://www.gov.uk/government/statistics/national-travel-survey-2014.

[66] P.-G. de Gennes, F. Brochard-Wyart, and D. Quere, *Capillarity and Wetting Phenomena* (Springer-Verlag, New York, 2004).

[67] C. Upton, *The Importance of Being Janus: Midland Speakers and the "North-South Divide"*, in *Middle and Modern English Corpus Linguistics*, edited by M. Markus, Y. Iyeiri, R. Heuberger, and E. Chamson (John Benjamins Publishing Company, Amsterdam, 2012), pp. 257–268.

[68] W. J. Heeringa, *Measuring Dialect Pronunciation Differences Using Levenshtein Distance* (University of Groningen, Groningen, 2004).

[69] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data* (Prentice-Hall, Englewood Cliffs, NJ, 1988).

[70] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik, *Cluster: Cluster Analysis Basics and Extensions. 2016. R package version 2.0.5. For new features, see the "Changelog" file* (in the package source), https://cran.r-project.org/web/packages/cluster.

[71] B Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis* (Wiley, Chichester, England, 2011).

[72] H Kuhn, *The Hungarian Method for the Assignment Problem*, Naval research logistics quarterly **2**, 83 (1955).

[73] S. N. Chiu, D. Stoyan, W. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications* (Wiley, Chichester, England, 2013).

[74] W. M. Rand, *Objective Criteria for the Evaluation of Clustering Methods*, J. Am. Stat. Assoc. **66**, 846 (1971).

[75] L. Hubert, *Comparing Partitions*, J. Classif. **2**, 193 (1985).

[76] W. Heeringa, J. Nerbonne, and P. Kleiweg, *Validating Dialect Comparison Methods*, in *Classification, Automation, and New Media* (Springer-Verlag, Berlin, 2002), pp. 445–452.

[77] P. Sammallahti, *The Saami Languages. An Introduction* (Dawi Girji, Karasjok, Norway, 1998).

[78] D. R. Preston, *Handbook of Perceptual Dialectology* (John Benjamins, Amsterdam, 1999).

[79] M. R. Spruit, W. Heeringa, and J. Nerbonne, *Associations among Linguistic Levels*, Lingua **119**, 1624 (2009).

[80] C. Gooskens, *Traveling Time as a Predictor of Linguistic Distance*, Dialectologia Geolinguistica **13**, 38 (2005).

[81] S. Montemagni, *The Space of Tuscan Dialectal Variation. A Correlation Study*, Int. J. Humanit. Arts Comput. **2**, 135 (2008).

[82] J. Nerbonne and P. Kleiweg, *Lexical Distance in Lamsas*, Comput. Humanit. **37**, 339 (2003).

[83] S. M. Allen and J. W. Cahn, *A Microscopic Theory for Antiphase Boundary Motion and Its Application to Antiphase Domain Coarsening*, Acta Metall. **27**, 1085 (1979).

[84] T. Ohta, D. Jasnow, and K. Kawasaki, *Universal Scaling in the Motions of Random Interfaces*, Phys. Rev. Lett. **49**, 1223 (1982).

[85] Y. Oono and S. Puri, *Large Wave Number Features of Form Factors for Phase Transition Kinetics*, Mod. Phys. Lett. B **02**, 861 (1988).

[86] S. V. Patankar, *Numerical Heat Transfer and Fluid Flow* (McGraw-Hill, New York, 1980).

[87] T. Leinonen, *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects* (Rijksuniversiteit Groningen, 2010).

[88] David Morin, *Introduction to Classical Mechanics* (Cambridge University Press, Cambridge, England, 2008).

[89] *NIST Handbook of Mathematical Functions*, edited by F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (Cambridge University Press, Cambridge, England, 2010).

[90] D. Brockmann, L. Hufnagel, and T. Geisel, *The Scaling Laws of Human Travel*, Nature (London) **439**, 462 (2006).

[91] M. C. Gonzlez, C. A. Hidalgo, and L. A. Barabsi, *Understanding Individual Human Mobility Patterns*, Nature (London) **453**, 779 (2008).

[92] H. Hayakawa, T Ishihara, K. Kawanishi, and T. S. Kobayakawa, *Phase-Ordering Kinetics in Nonconserved Scalar Systems with Long-Range Interactions*, Phys. Rev. E **48**, 4257 (1993).

[93] T. Blanchard, M. Picco, and M. A. Rajabpour, *Influence of Long-Range Interactions on the Critical Behavior of the Ising Model*, Europhys. Lett. **101**, 56003 (2013).

[94] B. D. Joseph and R. D. Janda, *The Handbook of Historical Linguistics* (Blackwell, Oxford, 2003).

[95] C.-M. Pop and E. Frey, *Language Change in a Multiple Group Society*, Phys. Rev. E **88**, 022814 (2013).

[96] P. Roach, *English Phonetics and Phonology* (Cambridge University Press, Cambridge, England, 2009).

[97] M. Davenport and S. J. Hannahs, *Introducing Phonetics and Phonology* (Routledge, Oxford, 2010).

[98] Leo Tolstoy, *War and Peace* (Penguin, Toronto, 1869).

[99] P. M. B. Vitányi, *Tolstoy's Mathematics in War and Peace*, Math. Intelligence **35**, 71 (2013).

[100] J. K. Chambers, *The Canada US Border as a Vanishing Isogloss: The Evidence of Chesterfield*, J. Engl. Ling. **23**, 155 (1995).

[101] G. Bailey, T. Wilke, J. Tillery, and L. Sand, *The Apparent Time Construct*, Lang. Var. Change **3**, 241 (1991).

[102] G. Baxter and W. Croft, *Modeling Language Change Across the Lifespan: Individual Trajectories in Community Change*, Lang. Var. Change **28**, 129 (2016).