

Using spatial patterns of English folk speech to infer the universality class of linguistic copying

James Burridge ^{*}*School of Mathematics and Physics, University of Portsmouth, Portsmouth PO1 3HF, United Kingdom*Tamsin Blaxter [†]*Gonville & Caius College, Cambridge CB2 1TA, United Kingdom*

(Received 15 May 2020; accepted 16 September 2020; published 14 October 2020)

Both linguistic and genetic evolution involve copying and mutation of variants. The simplest copying process assumes that variants are reproduced at a rate equal to their current frequency, exemplified by Kimura's stepping stone model of neutral evolution, and the voter model. In this case, spatial patterns are driven by noise. In the linguistic context, an alternative possibility is that speakers preferentially select variants which are already popular, yielding patterns driven by surface tension, exemplified by the Ising model. In this paper, we model language change using a spatial network of speakers, inspired by the Hopfield neural network. The model's universality class—Voter or Ising—is determined by speakers' learning function. We view maps generated by the Survey of English Dialects as samples from our network. Maximum likelihood analysis, and comparison of spatial auto-correlations between real and simulated maps, indicates that the underlying copying processes is more likely to belong to the conformity-driven Ising class.

DOI: [10.1103/PhysRevResearch.2.043053](https://doi.org/10.1103/PhysRevResearch.2.043053)

I. INTRODUCTION

Languages are complex, constantly evolving structures which take a wide variety of forms [1,2]. The processes of changing form and structure are referred to as *language change*, and involve evolutionary processes which can vary substantially between different parts of the linguistic system (sounds, word structures, syntax and vocabulary); changes may be driven by purely linguistic effects, social phenomena, migration, geography, technology and changes in wider society [3–6]. Nevertheless, every language is generated and maintained by a large number of interacting speakers with similar properties (vocal apparatus, a need to communicate, to display status and cooperate). It is therefore natural for statistical physicists to construct models which capture how languages arise and evolve, based on the interactions of such agents [7–15]. We provide a glossary of important linguistic terms in Appendix F.

A. Two kinds of copying

The neutral model of molecular evolution [16–19] assumes that every copy of every allele of a gene in the current generation of an organism is equally likely to be copied into the next.

Therefore the probability that an organism in the next generation will have a given allele is equal to the relative frequency of that allele in the generation before it. The linguistic analog of a gene is a *linguistic variable* and the analog of an allele is a *linguistic variant*. Variants are different speech forms which play the same role in the language. These might be different words for the same object or idea, systematic differences in pronunciations, or alternative grammatical rules. An example is the word for the prickly mammal most commonly known as a hedgehog. Historically, variants included *urchin*, *pricky back urchin*, *hedgehog*, and *hedgeboar*. A complete list of variables and their variants used in our study is given in Appendix E.

Suppose that we think of language evolution as an iterative process whereby speakers select or copy linguistic variants from those currently in use, or invent new ones. The linguistic analog of neutral molecular evolution is then that variants are selected by speakers with a probability equal to their current relative frequency. Neutral molecular evolution was introduced by Motoo Kimura [19] who analysed the spatial genetic variations which it generated using his *Stepping stone model* [18]. In physics and applied probability, this is the Voter model [20,21]. In linguistics, the Utterance selection model [13] is an example of a copying process which has correspondences with neutral evolution (Wright Fisher diffusion [22]).

An important property of evolutionary processes where organisms adopt or receive variants with probabilities equal to their current relative frequencies, is that changes are driven by noise alone. Despite its simplicity, and the complexity of real languages, such noise driven evolution remains a surprisingly robust *null* model of language change [23,24]. Although noise (diffusion) driven evolution, and “neutral evolution” are often used synonymously [23,24], a more general definition of

^{*}james.burridge@port.ac.uk[†]ttb26@cam.ac.uk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

neutral evolution is any copying process which is *symmetric* with respect to variants. This means that no variant has intrinsic advantage. To avoid confusion, we will refer to the case where selection probabilities are given by current (or observed, or recalled) relative frequencies as *proportional copying*, rather than neutral evolution.

Recent work on the spatial evolution of language (in both birds [25] and humans [10,26,27]) suggests that geographical boundaries between language features, known to linguists as *isoglosses*, may be analogous to the domain walls seen in classical lattice models of statistical physics [20,28] undergoing surface tension-driven coarsening [29]. A well known example of this coarsening process is exhibited by the Ising model evolving according to Glauber dynamics [30]. The surface tension effect requires some nonlinearity in the local copying rule [31,32], which in the social context implies a form of social conformity or majority rule [33–35]. From an individual perspective, conformity of this kind is beneficial if there is an advantage to matching the speech patterns of those with whom you interact. Together with the existence of isoglosses, this would appear to provide evidence against proportional copying. However, noise driven evolution can also generate distinct spatial domains. For example, the voter model [36], in which agents select their state by copying a randomly selected neighbor [20], is a form of proportional copying, because states are reproduced with a probability equal to their local frequency. Although the voter model lacks surface tension [3] it still evolves towards increased spatial order, characterized by logarithmically decaying correlations, and interfaces driven by noise. This raises the question: in cases where local speaker to speaker copying is responsible for geographical variations in language use, is proportional copying sufficient to explain observed patterns, or is a nonlinear copying rule more likely?

B. Purpose of the paper

The primary purpose of this paper is to address the above question, and to do this we define a model of language evolution in which speakers’ current linguistic behavior is related to their observations of past community via an *activation* or learning function. The deterministic version of our model is equivalent to a Hopfield neural network [37,38]. Hopfield networks represent a particularly simple model of agents who learn or copy from the previous behavior of others. Similar systems of dynamical equations, in which the rate of change of state is equal to the difference of the output of some learning or selection process, and the current state, form the basis of a number of language models [10,11,13,39–41]. The discrete stochastic Hopfield network may also be viewed as a generalized dynamical version of the Ising model [42].

We investigate two kinds of activation which, if the network is embedded in two dimensional space, produce behavior analogous to the classical voter model (proportional activation) or Ising model (conformity driven activation). The true dynamics of language evolution, if it can be captured mathematically at all, is certainly more complicated than our simple model. However, our assumption is that, as in physical systems, at large scales (national linguistic surveys) many small scale details of the evolutionary process become

irrelevant. In physics, systems driven by proportional and conformity driven copying belong to different universality classes with different spatial coarsening dynamics [20,29,31]. In particular, conformity leads to partial predictability of spatial language distributions [10,27].

Our data source is the Survey of English Dialects (SED) [43], a large scale survey of disappearing traditional English rural “folk-speech,” carried out in the 1950s. The relative lack of mobility, compared to modern people, of the communities within which the language features recorded by the SED evolved over the preceding centuries allows us to ignore migration in our coarse grained dynamics, which is then driven only by local copying. This allows the copying rule to be tested under controlled conditions, analogous to those found in classical lattice models. While modern surveys have generated vastly more data [44,45], the spatial patterns of language features have been mixed and diluted by movement and connectivity. The question of whether proportional copying provides an adequate description of language evolution in a simpler age is a fundamental one: the answer can inform the construction of more sophisticated models of the modern world, where the factors affecting language evolution are more diverse and difficult to model.

C. Structure of the paper

The structure of the paper is as follows. In Sec. II, we introduce our discrete time stochastic language model, and provide a continuous time approximation. In Sec. III, we explore in detail the spatial behavior of the model in the case of conformity driven copying, and proportional copying. In Sec. IV, we use spatial correlations, and maximum likelihood methods to infer which copying rule is more likely to be responsible for the spatial patterns of language use observed in English folk speech.

II. LANGUAGE COMMUNITY MODEL

We require a simple model of language change which can incorporate both proportional (noise driven) copying and conformity driven evolution. We begin by introducing a deterministic language model of a network of speakers, which may be seen as an adaptation of the continuous Hopfield network [37,38,42], which has its origins in the Ising model [42]. We then introduce stochasticity which will allow us to explore noise driven evolution.

A. Hopfield-like model

Consider a linguistic variable with $q \in \mathbb{N}$ variants, and let $v_{ik}(t)$ be the relative frequency with which speaker i uses variant k at time t . We write $\mathbf{v}_i = (v_{i1}, v_{i2} \dots, v_{iq})$ for the vector such frequencies. Speakers learn by listening to, and copying, others. This vector $\mathbf{v}_i(t)$ must therefore depend on the past behavior of other speakers, weighted by their influence on speaker i . A simple model of this information is the following time integral, capturing speaker *perceptions* or *memory*

$$\mathbf{u}_i(t) \triangleq \frac{1}{\tau_m} \int_{-\infty}^t e^{\frac{s-t}{\tau_m}} \sum_j \omega_{ij} \mathbf{v}_j(s) ds \quad (1)$$

where \triangleq denotes a definition. Here, τ_m , the *memory length*, specifies the typical time for which information remains important to current behavior. The number $\omega_{ij} \in [0, 1]$ is the influence of speaker j on speaker i , and we assume that the total influence is one

$$\sum_j \omega_{ij} = 1. \quad (2)$$

This condition means that the memory belongs to the q -dimensional simplex Δ^q ; it is a relative frequency. The memory (1) may be viewed as a network version of the linguistic memory defined in Ref. [10], and is a deterministic analog of the stochastic memory in the Utterance selection model [13]. We define the relationship between the way that people speak, and their memory, using an *activation function* $\mathbf{g} : \Delta^q \rightarrow \Delta^q$

$$\mathbf{v}_i(t) = \mathbf{g}(\mathbf{u}_i(t)). \quad (3)$$

The activation encodes language learning and copying processes, and innovation, which generate change. Differentiating (1) with respect to time, we obtain

$$\tau_m \frac{d\mathbf{u}_i}{dt} = \sum_j \omega_{ij} \mathbf{g}(\mathbf{u}_j) - \mathbf{u}_i. \quad (4)$$

This is the defining equation for a continuous Hopfield network [38,42], generalized to the case of more than two states (variants). In its original form the network was a model for a collection of neurons which were *activated* by electrical pulses from other neurons, mediated by synaptic interconnection strengths ω_{ij} . In this context, the activation function modelled the pulse required to trigger the neuron to fire. Equations of the form (4)—the difference of the output (generally construed) of a learning process, and its input—appear in a number of different language models [10,11,13,39,40]. In the linguistic context, the activation function may also be referred to as the *learning function* or *copying rule*. We will use \mathbf{g} to specify our copying process.

B. Stochastic model

Language evolution is not deterministic, and if we wish to use our model to explore proportional copying ($\mathbf{g}(\mathbf{u}) = \mathbf{u}$), then we must introduce an element of stochasticity. One potentially unrealistic aspect of the deterministic Hopfield model (4), is that a speaker exposed to a wide range of variants in similar proportions may persist in using all of them interchangeably, rather than settling on one. This behavior is removed if we interpret \mathbf{v}_i as a probability mass function over the set of possible language states, which are reselected at discrete time intervals.

We introduce a second time constant $\tau_s \leq \tau_m$, the *switching time*, which determines the typical time required for a speaker to change their selection. At time t , the *emitted* state $\mathbf{X}_i(t)$ of speaker i is then a sample from the probability mass function $\mathbf{v}_i(t - \tau_s)$, so that

$$\mathbb{P}(\mathbf{X}_i(t) = \mathbf{e}_k) = v_k(t - \tau_s), \quad (5)$$

where \mathbf{e}_k is a unit vector in the direction of the k th linguistic state. Defining $\delta\mathbf{u}_i(t) \triangleq \mathbf{u}_i(t) - \mathbf{u}_i(t - \tau_s)$, we introduce the

following discrete analog of the Hopfield model (4)

$$\delta\mathbf{u}_i(t) = \frac{\tau_s}{\tau_m} \left(\sum_j \omega_{ij} \mathbf{X}_j(t) - \mathbf{u}_i(t - \tau_s) \right). \quad (6)$$

For given τ_m , as $\tau_s \rightarrow 0$, we retrieve (4). For larger τ_s , speakers select a state and then stick to it for longer, increasing the stochasticity in \mathbf{u}_i . For given τ_s , increasing τ_m slows down the deterministic component of the dynamics, and reduces stochasticity by averaging over a larger sample of random updates.

Implicit in our model (6) is the assumption that speakers are immortal. We assume that the effects of birth, ageing and death may be captured by the iterative updating process, with the values of τ_m and τ_s depending on the way in which speakers learn and adapt their linguistic behavior over time. Much historical linguistic [46] and sociolinguistic [47] work on language change is predicated on the assumption that language behavior is mainly acquired in childhood and changes little later on; among other things, this is based on the very well-established findings from second language acquisition that language-learning abilities are highly degraded in adults. If this is so, then the memory length τ_m will be of the order of one human lifetime and τ_s will be of a similar order. However, if speakers remain flexible throughout their lives, for which there is some evidence at least for certain variables [48,49], then τ_m would be relatively short (a few years). There is some evidence in the case of grammatical variants [50] that adults, when faced with multiple ways to say something, match the usage frequencies in the language they have been exposed to. In this case τ_s would be small. We will see in Sec. III B that in the case of proportional copying, when stochasticity matters, it is the ratio τ_s/τ_m which controls spatial patterns, not their absolute values, which we remain agnostic about. If noise is responsible for patterns in real linguistic distributions, then we will find that this ratio needs to be close to one.

1. Spatial version

To analyze the spatial behavior of our model we divide space into a grid of square cells with side a , each with population N . We write $\mathcal{C}(\mathbf{r})$ for the set of speakers in the cell centered on \mathbf{r} . The cell average memory is then

$$\mathbf{u}(\mathbf{r}) \triangleq \frac{1}{N} \sum_{i \in \mathcal{C}(\mathbf{r})} \mathbf{u}_i, \quad (7)$$

with $\mathbf{v}(\mathbf{r}) \triangleq \mathbf{g}(\mathbf{u}(\mathbf{r}))$. We have omitted time dependence for brevity. We let $\langle \mathbf{r} \rangle$ denote the set of cell centres which are nearest neighbors to \mathbf{r} and introduce the *interaction range*, σ , which measures the typical distance over which speakers are in contact. We assume that cell-aggregated interaction strengths satisfy

$$\sum_{i \in \mathcal{C}(\mathbf{r})} \omega_{ij} = \begin{cases} 1 - 2\left(\frac{\sigma}{a}\right)^2 & \text{if } j \in \mathcal{C}(\mathbf{r}) \\ \frac{1}{2}\left(\frac{\sigma}{a}\right)^2 & \text{if } j \in \mathcal{C}(\mathbf{r}'), \mathbf{r}' \in \langle \mathbf{r} \rangle \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

According to this, each speaker, j , exerts a total influence of $1 - 2(\sigma/a)^2$ on those speakers in their own cell with whom they are in contact, and $(\sigma/a)^2/2$ on such speakers in each nearest neighbor cell. We will see below that this definition is

consistent with a Gaussian spatial interaction kernel. We also define the cell-aggregated emitted state

$$\mathbf{Y}(\mathbf{r}) \triangleq \sum_{j \in \mathcal{C}(\mathbf{r})} \mathbf{X}_j. \tag{9}$$

The expectation of this random variable is $\approx N\mathbf{v}(\mathbf{r})$ and its distribution is approximately multinomial($N, \mathbf{v}(\mathbf{r})$), which is the distribution we use in simulations. Averaging the noise term of our discrete dynamics (6) over one cell, we obtain

$$\frac{1}{N} \sum_{i \in \mathcal{C}(\mathbf{r})} \sum_j \omega_{ij} \mathbf{X}_j = \frac{1}{N} \sum_j \left(\sum_{i \in \mathcal{C}(\mathbf{r})} \omega_{ij} \right) \mathbf{X}_j \tag{10}$$

$$= \left[1 - 2 \left(\frac{\sigma}{a} \right)^2 \right] \frac{\mathbf{Y}(\mathbf{r})}{N} + \frac{1}{2} \left(\frac{\sigma}{a} \right)^2 \sum_{\mathbf{r}' \in (\mathbf{r})} \frac{\mathbf{Y}(\mathbf{r}')}{N} \tag{11}$$

$$= \frac{1}{N} \left[\mathbf{Y}(\mathbf{r}) + \frac{\sigma^2}{2} \nabla^2 \mathbf{Y}(\mathbf{r}) \right], \tag{12}$$

where ∇^2 is the discrete spatial second derivative

$$\nabla^2 \mathbf{Y}(\mathbf{r}) \triangleq \frac{1}{a^2} \left(\sum_{\mathbf{r}' \in (\mathbf{r})} \mathbf{Y}(\mathbf{r}') - 4\mathbf{Y}(\mathbf{r}) \right) \tag{13}$$

making (12) a saddle point approximation [51] to a Gaussian spatial average. The cell averaged version of the spatial dynamics (6) may then be written

$$\delta \mathbf{u}(\mathbf{r}) = \frac{\tau_s}{\tau_m} \left(\frac{1}{N} \left[\mathbf{Y}(\mathbf{r}) + \frac{\sigma^2}{2} \nabla^2 \mathbf{Y}(\mathbf{r}) \right] - \mathbf{u}(\mathbf{r}) \right) \tag{14}$$

where time dependence has been suppressed for brevity.

2. Continuous time approximation

It is useful to introduce a continuous time approximation to the dynamics (14), which is generalization of Wright Fisher diffusion [22]. A full derivation is given in Appendix A. Let \mathbf{O}_v be an orthogonal matrix ($\mathbf{O}_v^T = \mathbf{O}_v^{-1}$) whose last column is \mathbf{v} , and \odot denote the Hadamard (elementwise) product, and $\odot^{\frac{1}{2}}$ the Hadamard square root. Let $\mathbf{W} = (W_1, \dots, W_{q-1}, 0)^T$ be a vector of standard Brownian motions. Then the trajectories of the following stochastic differential equation provide a continuous time approximation of those generated by (14)

$$d\mathbf{u} = \frac{1}{\tau_m} \left(\mathbf{v} - \mathbf{u} + \frac{\sigma^2}{2} \nabla^2 \mathbf{v} \right) dt + \frac{\sqrt{\tau_s}}{\tau_m \sqrt{N}} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v d\mathbf{W}. \tag{15}$$

Here, \mathbf{r} dependence has been omitted for brevity. Equation (15) is discrete in space and continuous in time. Assuming that changes in the continuous version of \mathbf{u} are small over the interval $[t, t + \tau_s]$ then the discrete form (14) may be retrieved by integrating (15) from t to $t + \tau_s$. In the limit $\tau_s \rightarrow 0$, the noise term in (15) vanishes, as expected.

Approximating our evolution equation using (15) allows us to explore the effects of rescaling time units by a factor of

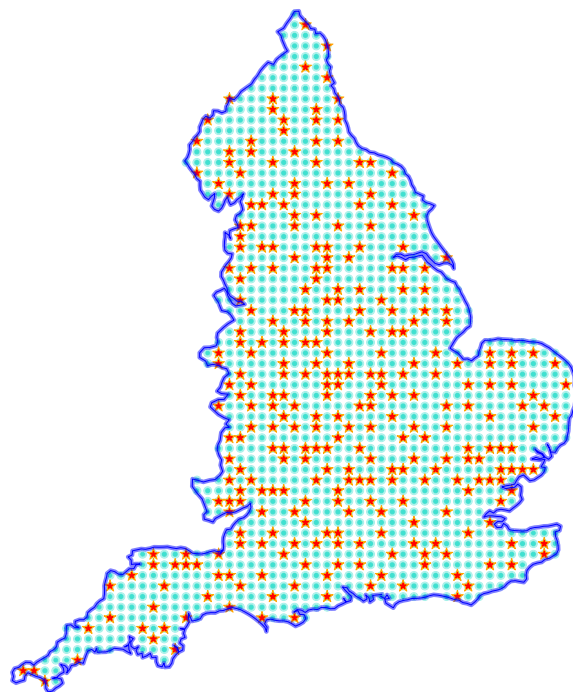


FIG. 1. Grid used for simulating English language features. Each light blue dot is the center of a 10 km \times 10 km grid square. Each red star is an SED survey location. There are 1329 grid squares, and 310 survey locations.

$c > 0$, that is ($\delta t_{\text{new}} = c\delta t_{\text{old}}$). This yields

$$d\mathbf{u} = \frac{1}{c\tau_m} \left(\mathbf{v} - \mathbf{u} + \frac{\sigma^2}{2} \nabla^2 \mathbf{v} \right) dt + \frac{\sqrt{\tau_s}}{\tau_m \sqrt{cN}} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v d\mathbf{W}. \tag{16}$$

Consider proportional copying, where $\mathbf{u} = \mathbf{v}$, with population N per cell and interaction range σ . If we reduce the population density to $N' = cN$, where $c < 1$, then provided we also increase the interaction range to

$$\sigma' = \frac{\sigma}{\sqrt{c}} \tag{17}$$

then the dynamics (16) is identical apart from a rescaling of time. Conversely, if we use simulated population $N_{\text{sim}} = cN$ and interaction range σ_{sim} then we will obtain spatial distributions with approximately the same statistical properties as if we have simulated the model with the full population, N , and interaction range

$$\sigma_{\text{eff}} = \sqrt{\frac{N_{\text{sim}}}{N}} \sigma_{\text{sim}}. \tag{18}$$

This will allow us to explore proportional copying by simulating smaller cell populations, which converge within a computationally feasible time frame. In this paper, our focus will be on English folk speech, as recorded in the Survey of English Dialects [43]. For simulations we divide England into a grid of 10 km \times 10 km squares, as shown in Fig. 1. There are 1329 grid squares, each containing 10^4 speakers giving a total population of 13.29 million, a level reached in the mid 1830s [52].

III. INTERFACES AND MATCHING

Our evolution equation (14) can generate spatial distributions where domains emerge in which a particular linguistic feature dominates. The structure and dynamics of these domains depends on the activation function \mathbf{g} . In this section, we explore the model's spatial behavior in the cases of proportional copying, $\mathbf{g}(\mathbf{u}) = \mathbf{u}$, and conformity driven copying. Understanding these spatial distributions will later allow us to infer which form of copying is more consistent with survey data.

A. Conformity driven copying

1. Conformity driven interfaces

For simplicity we consider a binary variable, so \mathbf{v} may be written

$$\mathbf{v} = (v, 1 - v). \quad (19)$$

We define the following activation function, which gives the probability for selecting variant one,

$$g(u) \triangleq \frac{e^{\beta u}}{e^{\beta u} + e^{\beta(1-u)}}. \quad (20)$$

The parameter β , which we call the *conformity number*, is analogous to inverse temperature in physical systems [20]. As $\beta \rightarrow 0$, corresponding to a very noisy or “hot” system, $g(u) \rightarrow \frac{1}{2}$ meaning that variants are selected entirely at random. As $\beta \rightarrow \infty$ speakers select the variant which is most common in their memory, leading to spatially ordered states. The fact that speakers using activation function (20) tend to adopt the behavior of the majority is the origin of the term *conformity driven* [33,35]. The critical inverse temperature $\beta_c = 2$ marks the transition between the disordered case, where variants persist in approximately equal proportions, and the ordered case when one variant dominates. These two cases correspond to the situations where $g(u) = u$ has one solution ($u = \frac{1}{2}$), and three solutions (two of which are stable), respectively.

An important property of conformity driven dynamics is its ability to maintain spatial interfaces [29,33]. To see how this occurs, consider an interface aligned along the y axis, so that $v = v(x)$. Assuming the population is large enough so that the system is well described by the deterministic component of its dynamics in (15), then the steady state shape of the interface, a smoothed step function, solves

$$\frac{\sigma^2}{2} v''(x) = \frac{1}{2} \left[1 + \frac{1}{\beta} \ln \left(\frac{v(x)}{1-v(x)} \right) \right] - v(x), \quad (21)$$

where $v''(x)$ denotes the lattice second derivative. The steepness of the interface at its midpoint may be found analytically (see Appendix C), and a simple measure of the interface width, ω , is the reciprocal of this gradient, which has asymptotic behavior

$$\omega(\sigma, \beta) \sim \sqrt{2}\sigma \left(1 + \frac{2 \ln 2}{\beta} \right) \text{ as } \beta \rightarrow \infty. \quad (22)$$

From this, we see that the interface width scales linearly with the interaction range, and becomes wider at higher temperatures. As $\beta \rightarrow 2^+$, the interface becomes infinitely wide as the

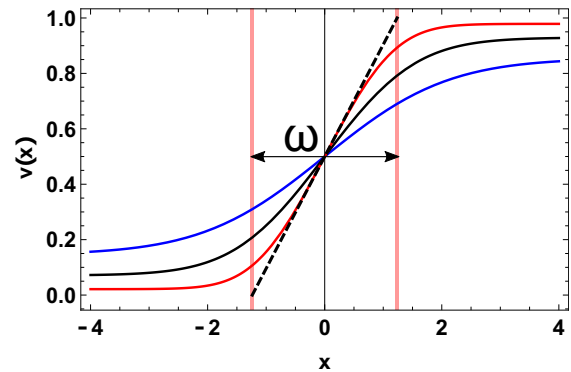


FIG. 2. Interface shape when $\sigma = 1$ and $\beta \in \{2.5, 3, 4\}$ (blue, black, red). Dashed line shows gradient of interface for $\beta = 4$, and red vertical lines show width of interface $\omega(\sigma, \beta)$ in this case.

system transitions to disorder. Figure 2 shows some example interfaces, obtained by numerically solving Eq. (21). Figure 3 shows how such interfaces can spontaneously form, starting from randomized initial conditions. This *coarsening* process is widely observed in two dimensional physical models of phase ordering, which have been adapted many times to model social phenomena, including language [8,10,14,15,25–27].

2. Matching probability functions

Differences between proportional and conformity driven copying affect the *matching probability*, $M(\mathbf{r}_1, \mathbf{r}_2)$, between two locations. This is the probability that the emitted states of

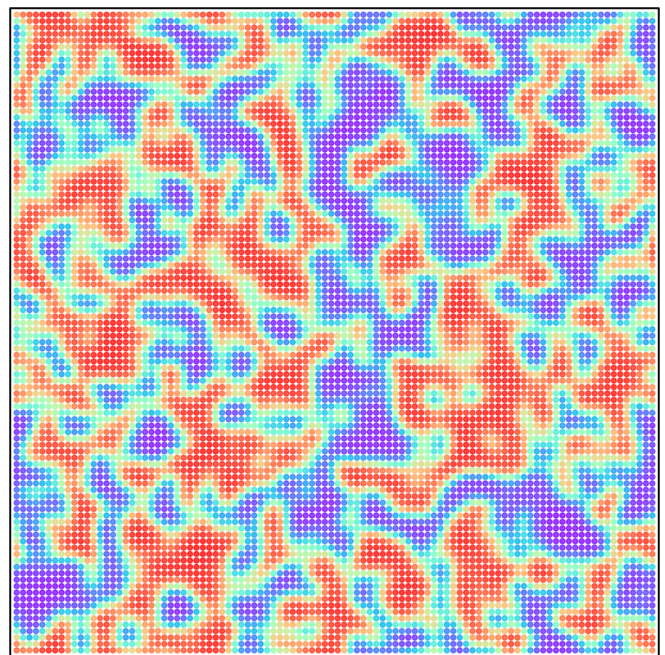


FIG. 3. Spatial distribution of the two state probability mass function $\mathbf{v}(\mathbf{r})$ over a 1000 km \times 1000 km toroidal system with $a = 10$ km. Parameter values $\sigma = 5$ km, $N = 10^4$, $\beta = 2.5$, and $\tau_m = 2$, $\tau_s = 1$. Evolution shown after 25 time steps starting from randomized initial conditions. Red and blue correspond to the two possible variants.

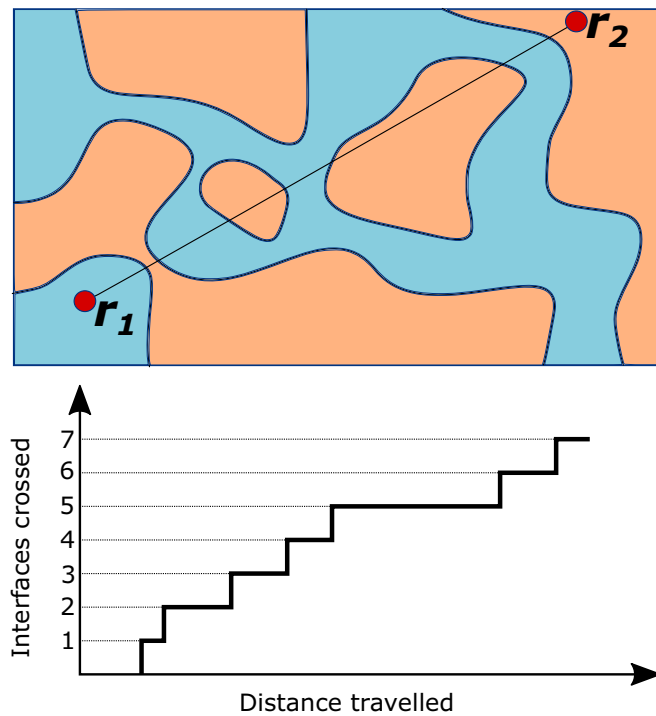


FIG. 4. Interface crossing count, traveling between \mathbf{r}_1 and \mathbf{r}_2 in a binary linguistic system. If an even number of interfaces are crossed, then the speakers at \mathbf{r}_1 and \mathbf{r}_2 will match with high probability.

two speakers in cells \mathbf{r}_1 and \mathbf{r}_2 are identical. In the case of a binary linguistic variable, we have

$$M(\mathbf{r}_1, \mathbf{r}_2) = 2v(\mathbf{r}_1)v(\mathbf{r}_2) - v(\mathbf{r}_1) - v(\mathbf{r}_2) + 1. \quad (23)$$

Matching probabilities (or correlation functions [20,29]) provide a simple means to characterize spatial distributions, and will be used in Sec. IV as part of our inference methodology. It is possible to estimate matching probabilities by direct simulation, or by adapting analytical techniques developed to calculate correlations in physical systems which exhibit phase ordering, starting from randomized initial conditions [29]. However, from a social-linguistic perspective these methods have some potential drawbacks. We know that the positions of interfaces can be influenced by initial conditions (determined by history and migration, and by the locations of innovations [53]), population distributions, geographical features, and localized cultural identities [2,3,10], and these may affect the sizes of domains. For this reason, as well as direct simulations, we also use an alternative approach in which typical domain size is a free parameter.

We imagine walking from \mathbf{r}_1 to \mathbf{r}_2 , and counting the interfaces crossed on the way (see Fig. 4). The crossing points of any straight line drawn across the system will form a point process [54]. To facilitate calculations, we will assume that the intervals between crossing points are independent random variables drawn from some distribution $f(\Delta r)$ (the marginal of the joint interval distribution), so that the locations of crossing points form a renewal process [54]. The simplest choice of marginal is exponential

$$f(\Delta r) = \frac{1}{\lambda} \exp\left(-\frac{\Delta r}{\lambda}\right), \quad (24)$$

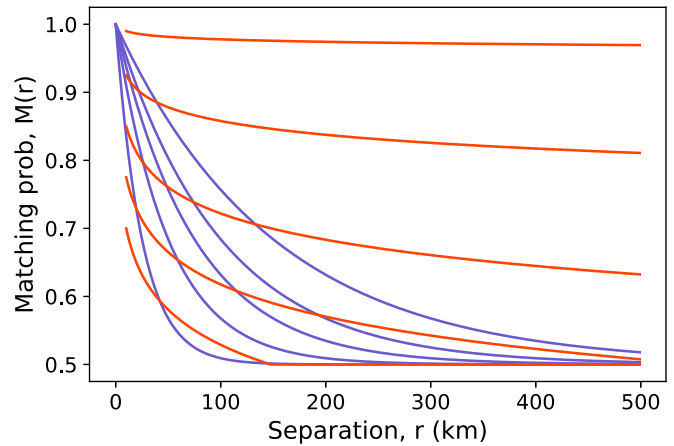


FIG. 5. Blue curves show conformity driven matching probability functions given by (27) for the five λ values given in Table I. Red curves show matching probabilities in the case of proportional copying given by (49) for five (b, c) values, also given in Table I.

where λ is the average distance between crossings—a measure of the typical size of a single domain. In this case the crossing points form a Poisson point process [54] with intensity λ^{-1} , and the number, $N(r)$, of crossings on a line of length r is a Poisson random variable with expectation $\mathbb{E}[N(r)] = r/\lambda$, and mass function

$$\mathbb{P}(N(r) = k) \triangleq p_k(r) = \frac{1}{k!} \left(\frac{r}{\lambda}\right)^k e^{-\frac{r}{\lambda}}. \quad (25)$$

Assuming that β is large, so that domains are linguistically *pure* and have interfaces which are narrow compared to domain size, then the matching probability for two points separated by a distance $r = |\mathbf{r}_1 - \mathbf{r}_2|$ is the probability that an even number of interfaces are crossed on the journey between them

$$M(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k=0}^{\infty} p_{2k}(r) \quad (26)$$

$$= e^{-\frac{r}{\lambda}} \cosh\left(\frac{r}{\lambda}\right). \quad (27)$$

These matching probabilities are plotted in Fig. 5. We note that exponentially decaying match probabilities (or correlations) are generic in phase ordering systems driven by short range interactions [29].

TABLE I. Parameter values for the five conformity driven (λ) and proportional copying (b, c) matching curves. Proportional copying (b, c) pairs estimated by fitting logarithmic function (49) with $\epsilon = 10$ km to simulated matching probabilities using model parameters $\sigma = 3.16$ km and $N = 5$ (approximately equivalent to $\sigma = 70$ m when $N = 10^4$), and $\tau_s = 1$, $\tau_m = 2$, $q = 2$.

Curve	1	2	3	4	5
Conformity, λ (km)	50	100	150	200	300
Proportional, b	0.700	0.775	0.850	0.925	0.990
Proportional, c	-0.0743	-0.0684	-0.0556	-0.0292	-0.0053

B. Proportional copying

Like conformity driven copying, proportional copying can also generate marked spatial variations. However, the driver of interface formation is noise rather than conformity, and the structure and dynamics of the interfaces, and the shapes of matching curves, are different.

1. Spatial variations from proportional copying

With proportional copying $\mathbf{v} = \mathbf{g}(\mathbf{u}) = \mathbf{u}$ our spatial system is driven by a combination of noise and spatial diffusion. To see this we decompose the cell-aggregated emitted state (9) into a deterministic and stochastic term $N^{-1}\mathbf{Y}(\mathbf{r}) \triangleq \mathbf{v}(\mathbf{r}) + N^{-1/2}\boldsymbol{\epsilon}(\mathbf{r})$, then our spatial dynamics (14) takes the form of a noisy diffusion equation

$$\delta\mathbf{v} = \frac{\tau_s}{\tau_m} \left(\underbrace{\frac{\sigma^2}{2}\nabla^2\mathbf{v}}_{\text{Diffusion}} + \underbrace{\frac{1}{\sqrt{N}}\left(1 + \frac{\sigma^2}{2}\nabla^2\right)\boldsymbol{\epsilon}}_{\text{Noise}} \right). \quad (28)$$

In the two variant case, the continuous time approximation (15) to this equation is

$$dv = \frac{1}{\tau_m} \frac{\sigma^2}{2} (\nabla^2 v) dt + \frac{\sqrt{\tau_s}}{\tau_m} \sqrt{\frac{1}{N} v(1-v)} dW. \quad (29)$$

An equation of this form, with different parameters, approximates Kimura's stepping stone model of neutral genetic evolution [18,21], where the population is divided into a grid of cells, termed "demes" or "colonies," each containing N_K individuals. Kimura's model [21] evolves in discrete time with each member of generation $n+1$ inheriting their type (one of two possible alleles A or B) from a member of generation n who is selected from the same cell with probability $1-m$ or from a nearest neighbor cell with probability m . A continuous time approximation for this process, is [22]

$$dv = \frac{m}{4} a^2 (\nabla^2 v) dt + \sqrt{\frac{1}{N_K} v(1-v)} dW, \quad (30)$$

where v is now the population fraction with allele A. The approximation applies when the population per cell is large enough so that the noise may be approximated with a Wright-Fisher diffusion [22]. Equations (28) and (30) are identical apart from their constant parameters. The spatial patterns generated by the stepping stone model have been studied in the equilibrium setting (see Refs. [21,22] and references therein), to model how genetic differences accumulate with distance.

Spatial variations in our model and the stepping stone model result from competition between diffusion and local noise. Whereas diffusion acts to equalize states between nearby sites, local noise generates spatial variations. For a sufficiently large interaction range, the diffusion term will equalize the linguistic/genetic state across the system much faster than noise effects can create locally distinct variations. The system behaves as a single well-mixed group (the "well mixed group" condition has recently been corrected [21] from its original form [55]). After some time, noise effects will drive the population into one of two pure states. When this occurs the system is said to have *fixed*. For spatially distinct domains to form, diffusion must act sufficiently slowly so that

parts of the system can temporarily enter different pure or near-pure states. Even then, the entire system will eventually fix in one or other state.

To see how interaction range affects this process, we give a simple derivation of the conditions under which distinct zones form. Consider a purely diffusive ($N \rightarrow \infty$) version of (29), which describes the spatial diffusion of particles, genes or linguistic variants with diffusion coefficient $D = \tau_s \sigma^2 / (2\tau_m)$. The root mean squared displacement of a diffusing particle, evolving according to (29) after time t , is

$$d(t) = \sqrt{4Dt} \quad (31)$$

$$= \sigma \sqrt{\frac{2t}{\tau_m}}. \quad (32)$$

This is the *diffusion distance*. In a system of linear size L , the time to diffuse across the system is then

$$t_{\text{mix}} \triangleq \frac{\tau_m}{2} \left(\frac{L}{\sigma} \right)^2. \quad (33)$$

We call this the *mixing time*. If diffusion acts sufficiently quickly, then the linguistic zone may be thought of as a single *panmictic* (random mixing) group [21,22] of size $M \approx N(L/a)^2$ obeying

$$dv = \frac{\sqrt{\tau_s}}{\tau_m \sqrt{M}} \sqrt{v(1-v)} dW_t. \quad (34)$$

The expectation of the time T required for this group to fix, starting from state $v(0) = x$ is

$$\mathbb{E}[T | v(0) = x] = -\frac{2M\tau_m^2}{\tau_s} \ln((1-x)^{-1-x} x^x). \quad (35)$$

The typical *fixation time*, starting from an equal proportion of each variant ($x = 1/2$) is then

$$t_{\text{fix}} \triangleq \frac{2 \ln 2 M \tau_m^2}{\tau_s}. \quad (36)$$

Now suppose that the mixing and fixing times are comparable. Setting $t_{\text{mix}} = t_{\text{fix}}$, we obtain the condition

$$4 \ln 2 \frac{\tau_m}{\tau_s} N \left(\frac{\sigma}{a} \right)^2 = 1. \quad (37)$$

In terms of population density $\rho = N/a^2$, this gives an approximate critical interaction range

$$\sigma_c \approx \sqrt{\frac{\tau_s}{\tau_m \rho}}, \quad (38)$$

where we have neglected the multiplicative constant $(4 \ln 2)^{-1/2} \approx 0.6$. If σ is of the order of σ_c or smaller, then variants cannot mix fast enough to keep the system in an effectively fully connected state. Subregions may form which are isolated for long enough to allow different pure states to form locally, before system wide fixation occurs. For the population to be panmictic (not geographical), σ must be substantially larger than σ_c [21]. A condition for "marked" spatial variation in the stepping stone model is given in Ref. [55]

$$mN_K < 1, \quad (39)$$

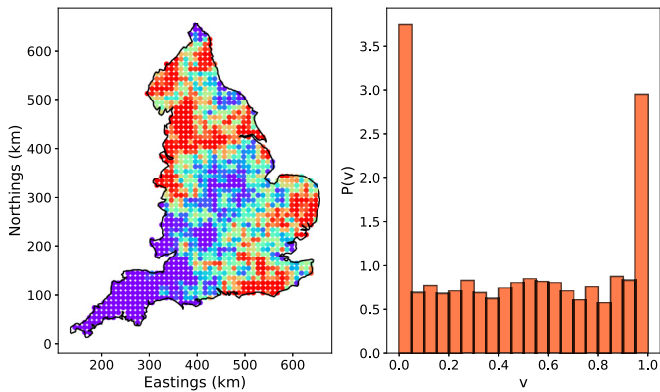


FIG. 6. External states of proportional copying $q = 2$ state system with $\tau_m = 2$, $\tau_s = 1$. The simulated population per cell and interaction range are $N_{\text{sim}} = 5$ and $\sigma_{\text{sim}} = 3.16$ km giving an effective interaction range of $\sigma_{\text{eff}} \approx 70$ m when $N = 10^4$ and $a = 10$ km. Histogram shows distribution of external states.

which is equivalent [by comparing (29) and (30)] to a critical interaction range

$$\hat{\sigma}_c = \frac{1}{\sqrt{2}} \sqrt{\frac{\tau_s}{\tau_m \rho}} \quad (40)$$

matching our heuristically derived value up to a constant close to unity.

Notice that σ_c does not depend on system size. As the local mixing rate increases, people become effectively more connected and the size of group which can be considered to have approximately the same memory state increases. This reduces the noise and slows the dynamics, meaning that a less rapid mixing rate is sufficient to keep even larger groups in the same memory state. Taking $\rho = 104.7$, which was the population density in England in 1841, we find that

$$\sigma_c \approx \frac{1}{10} \sqrt{\frac{\tau_s}{\tau_m}}. \quad (41)$$

Assuming that $\tau_s/\tau_m \approx 1$, then if the interaction range is substantially greater than a hundred meters, distinctive zones will not form. We know that, in fact, they do, and we assume that interaction ranges even smaller than this are unrealistic. Therefore, the only values of the ratio τ_s/τ_m which are consistent with proportional copying as a mechanism for spatial pattern formation, are close to one. We explore the range of such patterns by fixing the ratio $\tau_m/\tau_s = 1/2$ and varying the cell population (to change the effective interaction range) and the time for which simulations are run.

As noted in Sec. II B, using realistic cell populations and varying the interaction range requires unfeasibly long simulation times. We therefore simulate using a fixed, moderate interaction range and reduce the cell population, using relation (18) to estimate the effective interaction range which would generate similar spatial distributions if the cell population took a realistic value. The effect of reducing cell population/effective interaction range is illustrated in Figs. 6 and 7, which show the evolution of the proportional copying model with effective interaction ranges of 70 and 220 m. In the shorter range case, distinctive zones appear, creating a bimodal probability distribution of $v(\mathbf{r})$ over the system. When $\sigma_{\text{eff}} \approx 220$ m, although there are small spatial

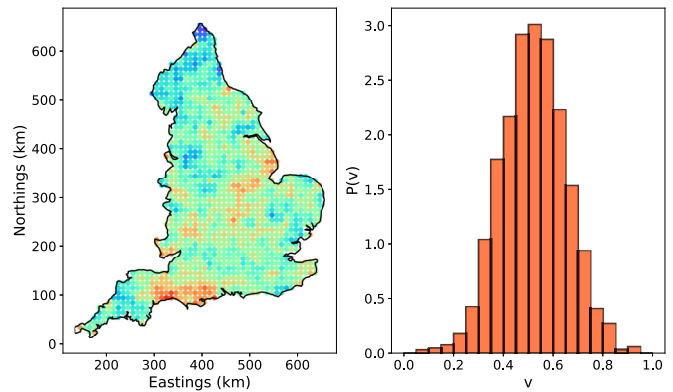


FIG. 7. External states of proportional copying $q = 2$ state system with $\tau_m = 2$, $\tau_s = 1$. The simulated population per cell and interaction range are $N_{\text{sim}} = 50$ and $\sigma_{\text{sim}} = 3.16$ km giving an effective interaction range of $\sigma_{\text{eff}} \approx 220$ m when $N = 10^4$ and $a = 10$ km. Histogram shows distribution of external states.

fluctuations, the distribution of $v(\mathbf{r})$ is clustered around a single value, meaning that the population as a whole are evolving as a single group. Interfaces and strong regional variations are therefore a feature of both proportional and conformity driven copying, but in the proportional case, for realistic population densities, very low geographical connectivity is required (even if τ_s/τ_m is close to one) for domains to form. In addition, from Fig. 6, the interfaces are geographically wide, and of a much more complex shape than in the conformity driven case, where their evolution is driven by a surface tension effect [10,20,29].

2. Matching curves with proportional copying

To calculate matching probabilities, we consider a very large (spatially invariant) system and derive an equation for $M(\mathbf{R}, \mathbf{R} + \mathbf{r})$ averaged over all locations \mathbf{R} to give a matching probability which depends only on displacement

$$m(\mathbf{r}) \triangleq \langle M(\mathbf{R}, \mathbf{R} + \mathbf{r}) \rangle_{\mathbf{R}}, \quad (42)$$

where $\langle \cdot \rangle_{\mathbf{R}}$ denotes the average over \mathbf{R} . Assuming that terms of order $\delta v(\mathbf{r}_1)\delta v(\mathbf{r}_2)$ can be neglected then the change in $M(\mathbf{R}, \mathbf{R} + \mathbf{r})$ per time step, making use of (28), when $\mathbf{r} \neq \mathbf{0}$, is

$$\begin{aligned} \delta M(\mathbf{R}, \mathbf{R} + \mathbf{r}) &= \sigma^2 \frac{\tau_s}{\tau_m} \left[v(\mathbf{R}) \nabla^2 v(\mathbf{R} + \mathbf{r}) + v(\mathbf{R} + \mathbf{r}) \nabla^2 v(\mathbf{R}) \right. \\ &\quad \left. - \frac{1}{2} (\nabla^2 v(\mathbf{R}) + \nabla^2 v(\mathbf{R} + \mathbf{r})) \right] + \text{noise} \\ &= \frac{\sigma^2}{2} \frac{\tau_s}{\tau_m} (\nabla_{\mathbf{R}}^2 + \nabla_{\mathbf{R}+\mathbf{r}}^2) M(\mathbf{R}, \mathbf{R} + \mathbf{r}) + \text{noise}. \end{aligned} \quad (43)$$

In the final line, we have made the position at which derivatives are taken explicit. Averaging over \mathbf{R} , we obtain, for $\mathbf{r} \neq \mathbf{r}$,

$$\langle \nabla_{\mathbf{R}}^2 M(\mathbf{R}, \mathbf{R} + \mathbf{r}) \rangle_{\mathbf{R}} = \nabla^2 m(\mathbf{r}), \quad (44)$$

$$\langle \nabla_{\mathbf{R}+\mathbf{r}}^2 M(\mathbf{R}, \mathbf{R} + \mathbf{r}) \rangle_{\mathbf{R}} = \nabla^2 m(\mathbf{r}), \quad (45)$$

$$\langle \text{noise} \rangle_{\mathbf{R}} \approx 0, \quad (46)$$

so

$$\delta m(\mathbf{r}) = \sigma^2 \frac{\tau_s}{\tau_m} \nabla^2 m(\mathbf{r}), \quad (47)$$

with continuous time approximation

$$\dot{m}(\mathbf{r}) = \frac{\sigma^2}{\tau_m} \nabla^2 m(\mathbf{r}). \quad (48)$$

Suppose that the interaction range is sufficiently small that locally pure spatial zones can form, and that the system starts from a spatially uncorrelated state. The typical size of pure zones will slowly grow, and a finite system will eventually consist of one single pure zone. However, there will typically be large fluctuations in the sizes and patterns of zones before this occurs. Within each pure zone the matching probability is one, so when such zones exist $m(\mathbf{0})$ will be close to one. Until fixation it will not equal one, because cells on boundaries of pure zones will be in a mixed state.

We obtain separation dependent matching probabilities from (47) using a quasistatic approximation, originally devised to calculate correlations in the voter model [20]. We let ϵ be a radius beyond which m is sufficiently slowly varying so that our discrete equation (48) may be approximated by a continuous diffusion equation. We then note that the continuous version of (48) describes the density of diffusing particles at a distance from a *source* region $\mathcal{S}(\epsilon) = \{\mathbf{r} \text{ s.t. } |\mathbf{r}| < \epsilon\}$, which appears once locally pure zones have formed, and is held at constant density equal to the value of $m(\mathbf{r})$ when $|\mathbf{r}| \approx \epsilon$. The density of particles at larger distances is initially $1/2$ (the matching probability between distant sites). Over time, particles from the source region will diffuse outwards, increasing the density away from the origin, corresponding to increased matching probabilities for larger \mathbf{r} values. At a time t after the initial formation of the source region, these extra particles have little effect on $m(\mathbf{r})$ beyond the diffusion distance $\sigma\sqrt{t/\tau_m}$, which divides the region around the source into a *near* and a *far* zone. If we assume that within the near zone the particle distribution has reached equilibrium then the solution within this zone is

$$m(\mathbf{r}) = b + c \ln\left(\frac{|\mathbf{r}|}{\epsilon}\right), \quad (49)$$

where b is the matching probability at short range, and c is fixed by the value of $m(\mathbf{r})$ at the edge of the near zone. Figure 8 shows that (49) can provide a good approximation to the empirical matching probabilities in the English dialect domain.

By fitting a large number of curves (49), we find a relationship, using kernel regression, between b and c , giving a one parameter family of correlation functions for the cases $N_{\text{sim}} \in \{5, 10, 50\}$ (Fig. 9). As $T \rightarrow \infty$, the system moves toward fixation, corresponding to $b \rightarrow 1$ and $c \rightarrow 0^-$. At earlier times, larger values of $|c|$ mean faster decaying matching probabilities, more spatial variations and higher spatial autocorrelation (see Sec. IV B). We will see in Sec. IV that the lowest cell population $N_{\text{sim}} = 5$ is best able to generate distributions consistent with the SED, so for the remainder of the paper we work with the family of matching curves generated for this case. The corresponding (b, c) values are listed in Table I, and the curves plotted in Fig. 5.

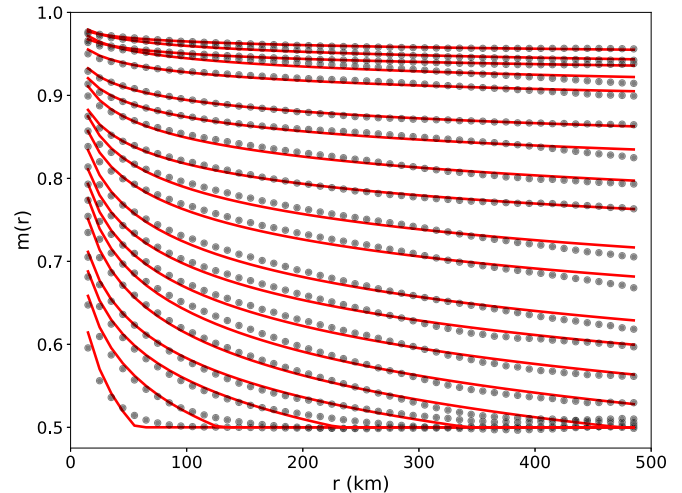


FIG. 8. Dots show equal time snapshots of the simulated correlation function $m(\mathbf{r})$ calculated from 500 realizations of proportional copying $q = 2$ state model with $\tau_m = 2$, $\tau_s = 1$, simulated on England starting from randomized, spatially uncorrelated initial conditions. Snapshots of the ensemble of systems were taken at a (quadratically increasing) sequence of 8 times in the interval $T \in [1, 2 \times 10^4]$. The simulated population per cell and interaction range are $N_{\text{sim}} = 5$ and $\sigma_{\text{sim}} = 3.16$ km giving an effective interaction range of $\sigma_{\text{eff}} \approx 70$ m when $N = 10^4$ and $a = 10$ km. Red curves show least squares fits to $\max(m(\mathbf{r}), 1/2)$ where $m(\mathbf{r})$ defined by (49) and $\epsilon = 10$ km.

3. Séguy's curve

Linguists may note that our matching function is closely related to Séguy's curve [56,57], which gives the relationship between geographical and linguistic distance. The linguistic distance between two locations may be defined as the number of variables that differ between them [58]. If all variables evolve according to the same dynamics then this is just $1 - M(\mathbf{r}_1, \mathbf{r}_2)$. In his original work Séguy fitted families of curves

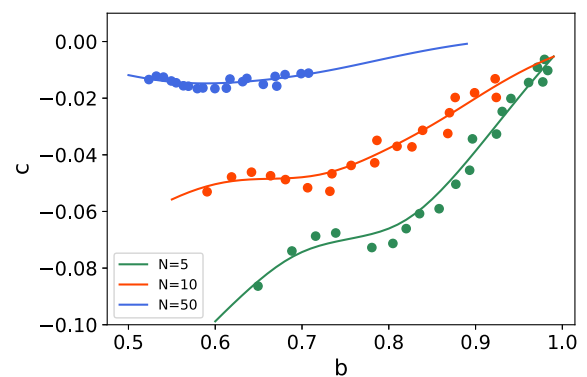


FIG. 9. Relationships between the parameters of the matching probability (49), with $\epsilon = 10$ km, obtained by fitting to snapshots at times in the interval $T \in [1, 2 \times 10^4]$ of the simulated correlation function $m(\mathbf{r})$ calculated from 500 realizations of the proportional copying model. In every case we have $\tau_m = 2$, $\tau_s = 1$, $\sigma_{\text{sim}} = 3.16$ km, $N = 10^4$ and $a = 10$ km. The three curves correspond to different simulated cell populations $N_{\text{sim}} \in \{5, 10, 50\}$. Solid lines show kernel density regression on simulated data with bandwidth $h = 0.05$.

based on logarithmic increase, generating “fat tailed” distance relationships more consistent with proportional copying than conformity driven dynamics. There is some debate as to what family of functions Séguy’s curve belongs to [58], and this question has recently been addressed from the point of view of Statistical Physics [10].

IV. INFERENCE

We now use data from the Survey of English Dialects (SED) to infer which class of copying rule provides a better description of the spatial distributions of language features which it recorded.

A. The Survey of English Dialects

The SED contains 310 survey locations within the British mainland, excluding the Isle of Wight and the Isle of Mann (see Fig. 1). The geographical distribution of language features recorded in the SED are the result of local copying processes with biases which depend on the linguistic feature in question, and on social factors. We view language change as a branching process which generates a single new variant at a time. In order to become established some mechanism must exist which, at least temporarily, biases speakers in favor of new variants. Such mechanisms might be socially conditioned (used to signal social group or generation) or linguistic (the new variant is innately preferable). However, the existence of long lived stable interfaces between variants suggests that such biases are in many cases weak, short lived or contextual. Geographical distributions will also be influenced by migration events and political changes. Our aim is only to understand the fundamental class of the copying process, assuming that variants are approximately equivalent. We wish to know if, in the absence of differences in intrinsic “fitness,” variants survive with a probability equal to their current frequency, or whether it is more probable that speakers preferentially select variants which are already more common.

We consider variables for which it is possible to identify a single branch in their evolution. In some cases this involves reducing multivariate maps to bivariate by merging or excluding variants. Where one variant transparently reflects an additional change modifying the output of an earlier change, we can merge these so that the dataset reflects only the distribution of the earlier change. Where the relationships among variants are nontransparent because a later change obscures an earlier one, or where a variant is formally unrelated to the others, there is no justification to merge it and we exclude the variant altogether. If the variant occurs only within an otherwise well-defined domain (Fig. 10), then we can exclude it and impute the missing data because firstly, we are modeling the changes in distribution of the other two variants and the boundary between the domains in which these are dominant remains clear; secondly, a distribution of this type suggests that the third variant is a later innovation within an existing region, even if we do not have specific historical evidence to support this. If, on the other hand, a problem variant is not embedded within a clearly defined domain but occurs along

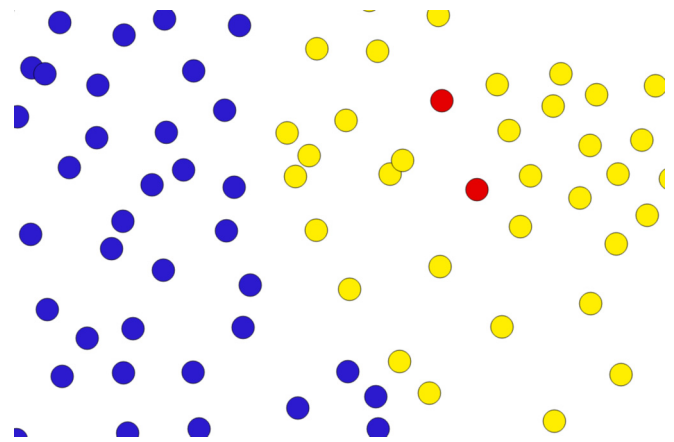


FIG. 10. A third variant embedded within a clearly defined domain.

another domain boundary (“isogloss”) (Fig. 11), there is no way to treat the variable as binary since we cannot reconstruct the distribution of the two variants we are interested in at a crucial point where they interact, and we lack evidence for the chronology of innovations. Variables falling into this class are excluded from our study. Figures 12 and 13 show binary distribution maps for 40 of the 68 variables considered. A detailed description of the linguistic variables used and, where relevant, their reduction to binary form, is given in Appendix E.

B. Spatial autocorrelation

For each survey map in our dataset, we wish to infer which of our two copying processes is more likely to have generated it. Inspection of the maps (Figs. 12 and 13) shows that while some variables exhibit well defined spatial interfaces, other maps are more disordered. Broadly speaking we expect conformity driven evolution to yield well defined domains with smooth boundaries, and proportional copying to generate a more complex pattern of spatial boundaries and disorder, if interaction range is short. For proportional copying with

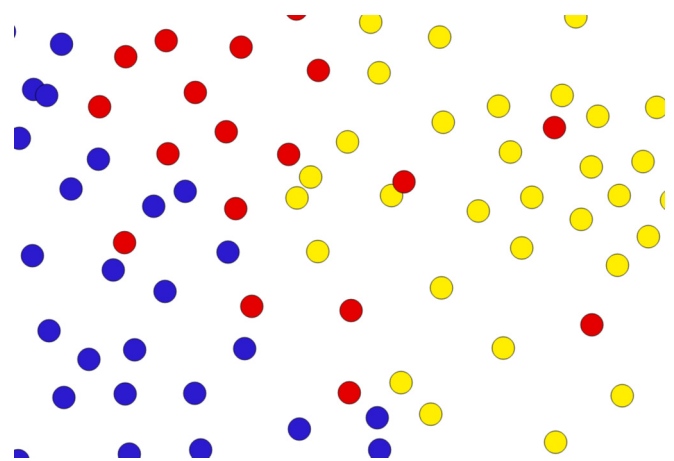


FIG. 11. A third variant not well-embedded within a clearly defined domain.

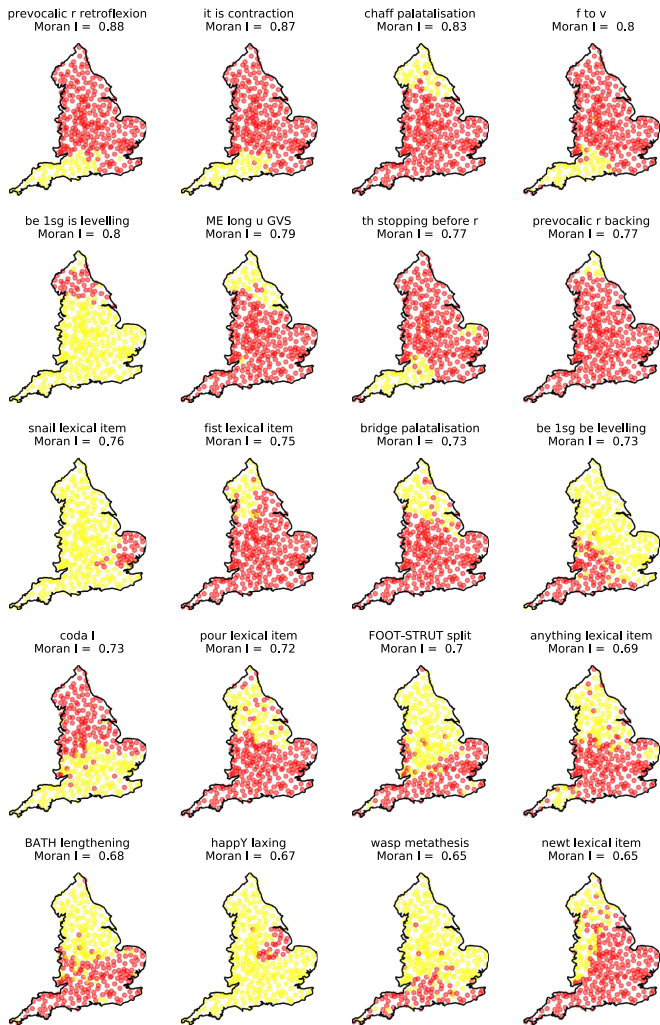


FIG. 12. SED maps of the twenty variables with highest Moran I . The title of each map gives the formal linguistic description of the variable, and the Moran I value using six nearest neighbors.

longer interaction ranges, we would expect limited spatial order in survey results unless system wide fixation has occurred. Because different parts of the linguistic system evolve by different processes and at different rates, and also carry different social messages, then we do not expect every map to be the result of the same underlying process. We therefore test each language feature individually.

1. Moran's I

We have characterized the spatial distributions generated by the different activation functions using spatial matching probabilities. However, these cannot be used to draw inference from individual maps, because the function $m(\mathbf{r})$ calculated by averaging over the locations in a single map will strongly depend on the particular distribution of that variable. We could only infer whether matching probabilities for that variable would exhibit exponential or logarithmic decay if we could rerun its history many times and average $m(\mathbf{r})$ over the results. To resolve this, in Sec. IVC, we use matching functions to generate approximate multivariate probability distributions

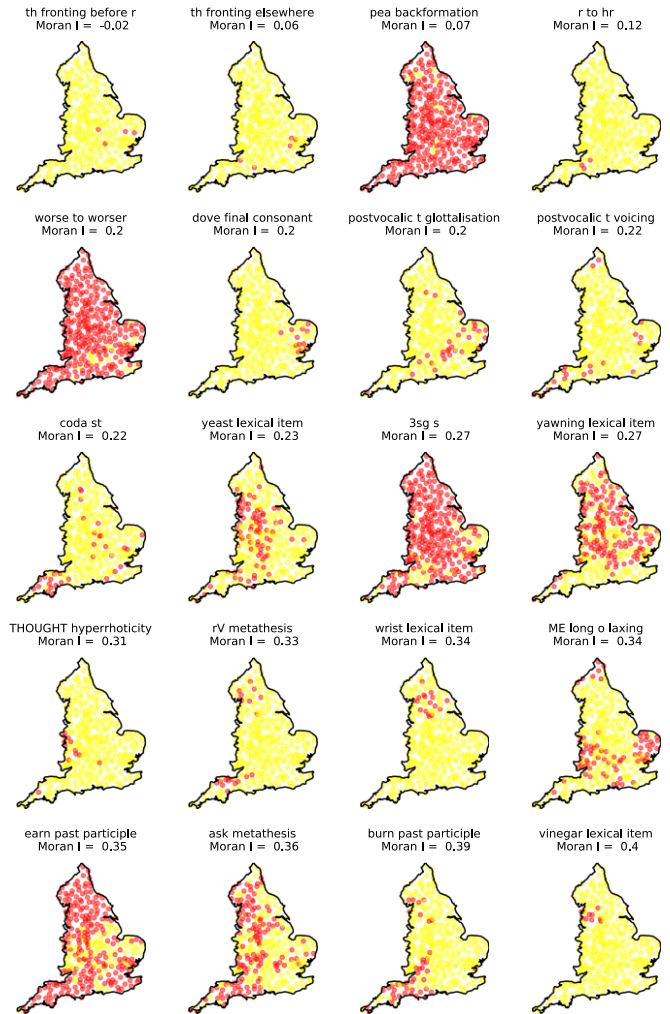


FIG. 13. SED maps of the twenty variables with lowest Moran I . The titles give the formal linguistic description of each variable, and the Moran I value calculated using six nearest neighbors.

over the set of all possible maps, allowing likelihood based inference. However, we first take a simpler approach based on the extent to which nearby locations are in the same state, known as *spatial autocorrelation*. A simple measure of this, previously used to study regional linguistic variation [59], is Moran's I [60], which, for N locations, is defined

$$I \triangleq \frac{N}{G} \frac{\sum_{ij} G_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}, \quad (50)$$

where x_i is the state of the i th location and G_{ij} a spatial weight associated with the pair (i, j) . The number G is the sum of all spatial weights. In our case the state at location i is the emitted state of the speaker selected for the language survey. Rather than represent this state using $\{\mathbf{e}_1, \mathbf{e}_2\}$, we use $x_i \in \{-1, 1\}$, so, if the survey contains an equal number of speakers using each variant then $\bar{x} = 0$. The spatial weights G_{ij} define what we mean by *nearby*. We take $G_{ij} = 1$ if survey location j is one of the six nearest neighbors of i and $G_{ij} = 0$ otherwise. The mean and standard deviation of the separation of locations with nonzero weights is then 22.5 ± 8.4 km.

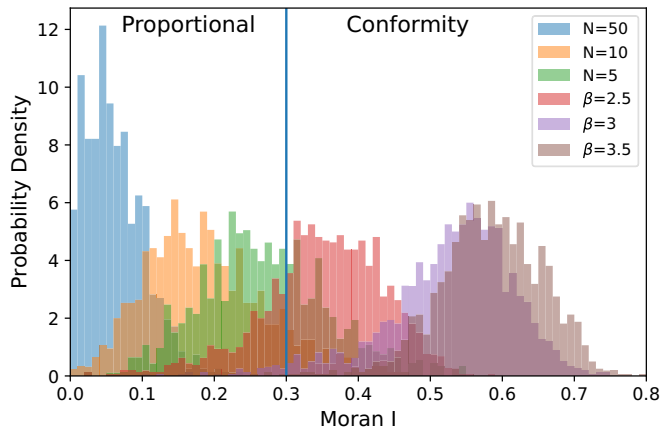


FIG. 14. Distributions of I values (six nearest neighbors) from direct simulation on England using $q = 2, \tau_m = 2, \tau_s = 1$. In the proportional copying case, we have $\sigma_{\text{sim}} = 3.16$ km, and $N_{\text{sim}} \in \{5, 10, 50\}$ giving effective interaction ranges $\sigma_{\text{eff}} \in \{70 \text{ m}, 100 \text{ m}, 220 \text{ m}\}$ when the true cell population is $N = 10^4$. In the conformity driven model we use $\sigma_{\text{sim}} = 5$ km, $N = 10^4$, and $\beta \in \{2.5, 3, 3.5\}$. Vertical line indicates where proportional/conformity model cases are more common.

Moran's $I \in [-1, 1]$ then measures the extent to which survey locations within this range match their state. Intuitively, if we have large single-variant domains with well defined, smooth interfaces then we expect high I value, because the regions of the system where miss-matches occur are one dimensional and maximally short (due to surface tension), and therefore occupy a small fraction of the total area. If, on the other hand, single variant domains have a complex boundary structure, or variants are otherwise widely dispersed, then we expect a low I value. This intuition is borne out by Figs. 12 and 13 which show the variables with, respectively, the highest and lowest I values.

2. Comparing simulations to the SED

To understand the relationship between our evolution models and Moran's I we directly simulate the English dialect domain starting from randomized initial conditions, and extract the emitted states of speakers at each SED survey location at a fixed sequence of time intervals. Using this data we then compute the I value of each sample using the same weights as the survey data, allowing comparison to the I values from the SED. The distributions of simulated I values are shown in Fig. 14. In the proportional copying case we vary the effective interaction range by changing the simulated population per cell, giving $\sigma_{\text{eff}} \in \{70 \text{ m}, 100 \text{ m}, 220 \text{ m}\}$. In the longer range case, spatial variations in external state are small, and most spatial variation is generated by the sampling process; we obtain I values which are close to zero. For the shorter effective interaction ranges, where linguistic subdomains are able to form, we obtain higher I values. Using the conformity driven activation function (20) we have two free parameters: the interaction range and the conformity number β . Whereas β determines the amount of noise in the bulk [32], both σ and β together determine the width of interfaces [Eq. (22)]. We set the simulated range to $\sigma_{\text{sim}} = 5$ km (note,

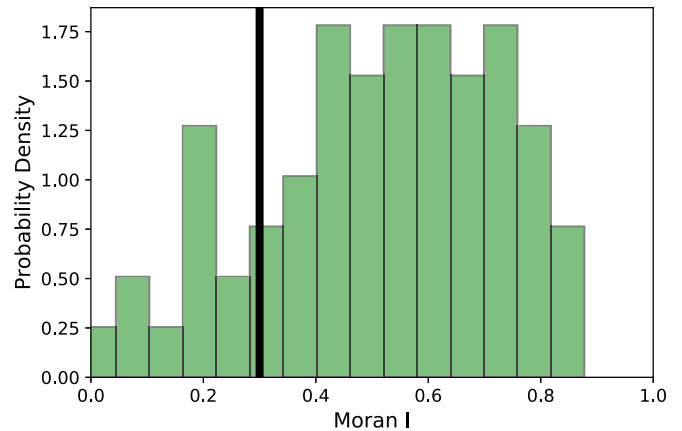


FIG. 15. Distributions of I values (six nearest neighbors) from the SED data. Vertical line indices where proportional/conformity model cases are more common in direct simulations. 82% of cases have $I > 0.3$.

villages recorded in the Domesday book are typically ≈ 2 km separated from their closest neighbor, with remarkable consistency between shires [61]). We then simulate the model for $\beta \in \{2.5, 3, 5\}$, noting that as $\beta \rightarrow 2$, the system approaches complete disorder, where variants exist in equal proportions in all locations.

Since our survey maps contain linguistically pure regions, we assume that values of β near the disorder transition are not a realistic model of linguistic behavior. The I distributions obtained from our three β values are shown in Fig. 14 and we see that they are substantially higher than the proportional copying values. For $I > 0.3$, the majority of samples are conformity driven. Figure 15 shows the distribution of I values computed from the SED. The majority (82%) of maps have I values which are more likely to have been generated by our conformity driven simulations. From here on we compare conformity driven evolution to the proportional copying model with lowest cell population $N_{\text{sim}} = 5$ on the basis that this model is most likely to be able to match realistic spatial distributions.

Moran's I is a simple and intuitive means to distinguish between different kinds of spatial distribution, and our analysis suggests that although the proportional copying model is capable of generating distributions which exhibit the kinds of spatial ordering seen in the SED, the majority of maps are more consistent with conformity driven evolution. However, Moran's I depends only on matching probabilities at close range when we know that in fact the differences between proportional and conformity driven copying are manifested in the full r -dependence of the matching probability function (Fig. 5). As an example of why this means that I -based inference may be problematic, we note that even if $\beta > 2$, so that interfaces exist, it is possible to create any desired level of short range spatial disorder by tuning β sufficiently close to $\beta_c = 2$. As noted above, such distributions are not attested in the data, but nevertheless have I values typical of the proportional model. We now consider an inference method which removes this issue by accounting for the full r dependence of matching probabilities.

C. Markov graphical models

To infer what model, and what parameters, are most likely to have generated a dataset, we require a *statistical model* of that data [62]. A minimum requirement is that the model provides a method for generating realizations of the data, given its parameters. Ideally the model will also provide the full probability distribution of the output of a single trial. In our case, a single trial corresponds to one SED survey map, and our simulations satisfy the minimum requirement of a statistical model of such maps. However, because the number of possible outcomes of each trial is a high dimensional random vector, no likelihood can be attached to a given map or set of maps—the sample space is too large. Moreover, as discussed in Sec. III A, when considering the possible arrangements of conformity driven interfaces, the natural parameter which describes matching probabilities is the density of interfaces in the system (or, equivalently, the average domain size), which cannot be directly controlled in simulations, and may depend on factors exogenous to our simple dynamics.

1. Pairwise Markov random field

An alternative statistical model which incorporates the full r dependence of matching probabilities, and gives the probability of any possible map, is the *Markov Random Field* or *Markov Graphical Model* [62,63]. This class of model began with the Ising model, and is now used in a range of fields including computer vision, spatial data analysis [64] and machine learning [63]. We have sufficient information about matching probabilities to calibrate a *pairwise model*, in which the probability of map $\mathbf{x} = (x_1, x_2, \dots)^T$, where $x_i \in \{-1, 1\}$, is

$$P(\mathbf{x}) = \frac{\exp(\frac{1}{2}\mathbf{x}^T\theta\mathbf{x})}{\mathcal{Z}(\theta)}, \quad (51)$$

where θ is a symmetric matrix of interaction strengths between all possible pairs of sites, and the normalizing constant \mathcal{Z} is the *partition function*. The probabilities assigned for different configurations \mathbf{x} do not depend on the diagonal elements of θ because $x_i x_i = 1$ for all i , so these elements contribute a multiplicative constant to the numerator of (51), which affects the value of the partition function. By convention we set $\theta_{ii} = 0$ for all i .

Samples from (51) may be obtained via Gibbs sampling [42]. Starting from a randomized initial state, we propose changes by selecting a single site i and setting $x_i = 1$ with probability [30]

$$p_i = \frac{1}{2} \left(1 + \tanh \sum_j \theta_{ij} x_j \right), \quad (52)$$

and $x_i = -1$ with probability $1 - p_i$. A distribution \mathbb{P} is the equilibrium of this update rule if the detailed balance condition is satisfied

$$\mathbb{P}(x_i = 1 \cap \mathbf{x}_{\setminus i})(1 - p_i) = \mathbb{P}(x_i = -1 \cap \mathbf{x}_{\setminus i})p_i, \quad (53)$$

where $\mathbf{x}_{\setminus i}$ denotes the states of all sites excluding i . That condition (53) is satisfied by (51) may be seen by noting that

$$\frac{P(x_i = 1 \cap \mathbf{x}_{\setminus i})}{P(x_i = -1 \cap \mathbf{x}_{\setminus i})} = e^{2\sum_j \theta_{ij} x_j} = \frac{p_i}{1 - p_i}. \quad (54)$$

2. Calibration

To determine the interaction matrix we use our exponential and logarithmic matching probabilities (27) and (49) (with parameters given in Table 1) to calculate the matching probability between every pair of nodes in the SED, based on their separations. For each matching curve, we obtain a matching probability matrix M_{ij} . The equivalent matrix for our statistical model $P(\mathbf{x})$ is given by

$$\hat{M}_{ij}(\theta) = \mathbb{E} \left[\frac{1 + x_i x_j}{2} \right], \quad (55)$$

which may be estimated by Gibbs sampling. We calibrate our model to the desired matrix by iteratively adjusting the interaction parameters using the descent rule [42]

$$\theta_{n+1} = \eta(M - \hat{M}(\theta_n)), \quad (56)$$

where $\eta > 0$ is a *learning rate*. That is, interaction strengths are incrementally increased or decreased to shift the model matching probabilities toward their targets. The practical (vectorized PYTHON) implementation of the method involves storing many independent realizations of the system (the random vector \mathbf{x}) where each element of each vector is initialized to ± 1 with equal probability, corresponding to $\theta_{ij} = 0$ for all i, j . After each iteration of (56), each vector is updated (using Gibbs sampling) a sufficient number of times so that the set of vectors $\{\mathbf{x}\}$ represent a sample from the current model, θ_n . The matching probabilities for this model are then estimated, and used to calculate θ_{n+1} .

We note an important difference between our calibrated model, which can, in principle, have interactions at all separations, and short range Ising-type models. The subcritical nearest neighbor Ising model, updated using Gibbs sampling, generates domains which grow larger over time leading to a progressively lower interface density [20,29]. In contrast, our model is calibrated so that, at least in the conformity driven case, the interface density stabilizes at a given target value (λ^{-1}). Examination of calibrated interaction matrices reveals negative interactions at ranges beyond the typical domain size, which limit the expansion of domains once the desired interface density has been achieved.

3. Inference

Having calibrated the interaction strengths, we can use our model to generate sample maps consistent with the target matching curves. A set of such maps are shown in Fig. 16. From these examples we see that the calibrated conformity driven model, as expected, generates well defined domains with smooth interfaces. As the density of interfaces declines, Moran's I increases. In the proportional copying case, although spatial domains appear, they are less well defined in early stage evolution, which is consistent with simulations of the external state shown in Fig. 6, and produces Moran I values which are typically lower than those obtained from interface-driven dynamics. By generating a much larger sample of maps from the two calibrated models, we can estimate the empirical distribution of proportional and conformity driven I values, as shown in Fig. 17, along with results for the SED. From this, we see that the SED and the conformity driven model generate a similar range and

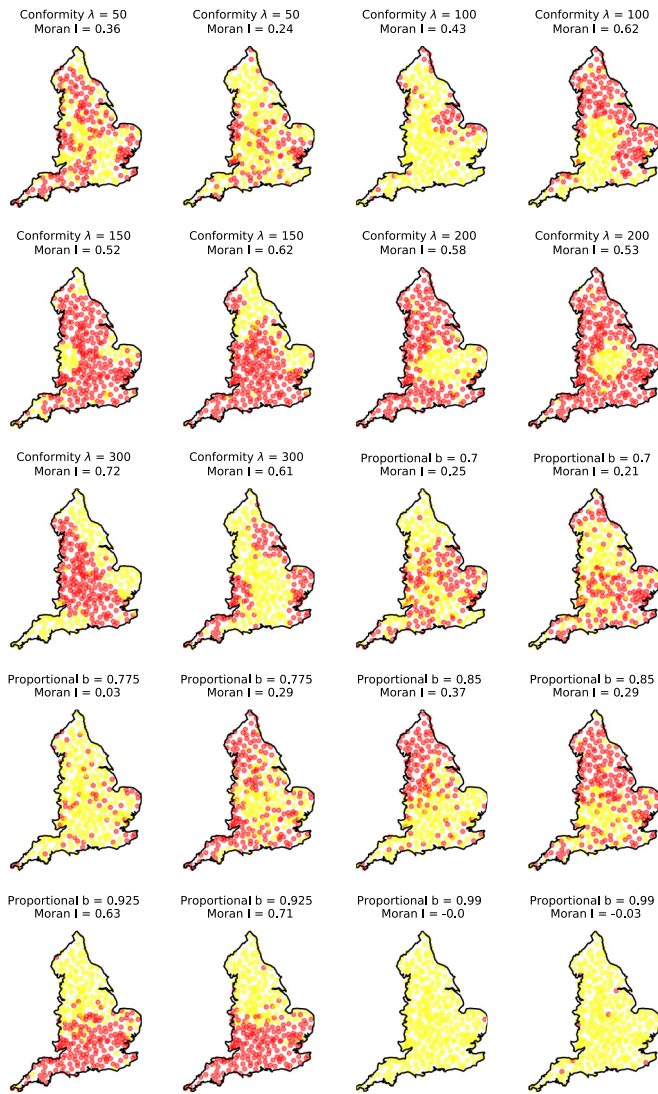


FIG. 16. Samples from Markov graphical models calibrated proportional copying and conformity driven matching curves in Fig. 5.

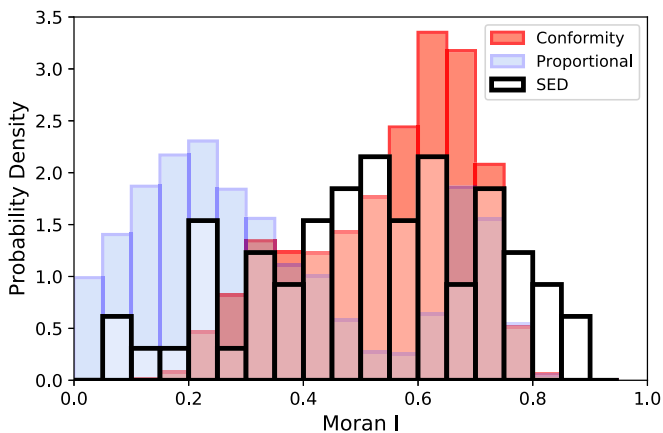


FIG. 17. Distributions of I values (six nearest neighbors) from the SED data, and from Markov random field models calibrated to matching functions described by the parameters in Table I and plotted in Fig. 5.

distribution of I values, with the proportional model tending to produce maps with lower values. We note that the distribution of proportional copying I values has a secondary peak around $I \approx 0.7$, which is not reproduced by direct simulation of the model (Fig. 14). This peak is produced by matching curve 4 in Table I, and highlights that fact that our statistical model is only an approximation to the true spatial distributions, for which a closed form probability distribution does not exist.

Beyond sampling individual maps we can also calculate the logarithmic probability of map \mathbf{x} , given the interactions θ calibrated to a matching curve

$$\ln P(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \theta \mathbf{x} - \mathcal{Z}(\theta). \quad (57)$$

To evaluate this expression we require an estimate for the partition function $\mathcal{Z}(\theta)$, which cannot be computed exactly due to the intractable sum over all possible states. We adopt the annealed importance sampling method, developed by Neal [65] (see Appendix D). For every map we can then estimate its likelihood for every matching curve to which we have calibrated θ . Of the 68 maps in our dataset we find 14 for which a proportional copying model matching curve has the highest likelihood. Therefore $\approx 80\%$ of maps are more likely to be conformity driven, consistent with the 82% result obtained using Moran's I . The mean domain size in conformity driven maps is $\bar{\lambda} = 182$ km with standard deviation 72 km.

We test our methodology by generating 100 samples for each calibrated model, and verifying that the average logarithmic probability of these samples is maximized for the model that generated them. The results are displayed in Fig. 18. In Fig. 18(a), typical logarithmic probabilities increase with λ because there are many more ways to cover a map with small domains than there are to cover it with large ones. Likewise, logarithmic probabilities of the proportional copying maps increase with b , which tends to one as fixation is approached and domains grow larger in size. With reference to Fig. 18(b), we note that the logarithmic probabilities span a larger range of values than then conformity driven case for $\lambda > 50$ km. This reflects the fact that the proportional copying model displays a broader range of behavior. The same model can generate maps with or without domains, and a single realization of the dynamical model can generate a wide variety of different spatial patterns before it reaches fixation.

V. DISCUSSION

Languages are complex structures which exhibit a wide range of change processes. These processes have been catalogued and studied by linguists for centuries, with the volume and intensity of research rapidly rising in the late twentieth and early twenty first century [2,3,66]. Research into language evolution has become increasingly quantitative [2,59] and interdisciplinary [7,10,13], with models inspired by statistical physics. The analogy between linguistic and genetic evolution is also long standing [23,67]. In neutral molecular evolution, every copy of every gene has an equal chance of surviving into the next generation, and a linguistic analog of this is proportional copying. Language models in this class have been remarkably successful in describing aspects of language change [13,23,24,68], perhaps in part because noise driven models display a rich variety of behavior.

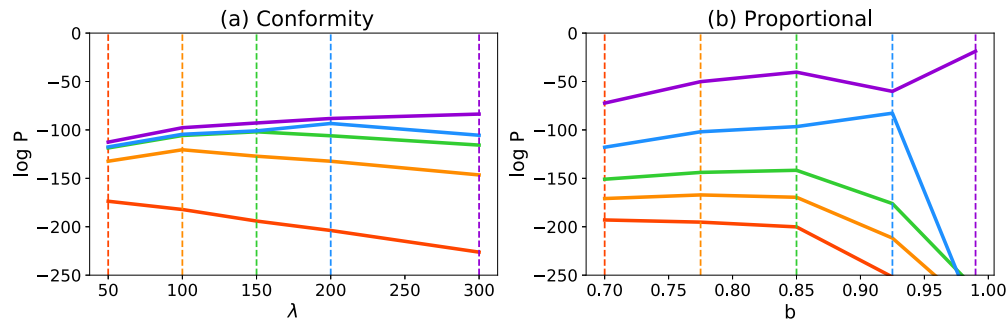


FIG. 18. (a) Each curve shows the average logarithmic probability (proportional to logarithmic likelihood) of 100 samples from a single conformity driven model, calibrated to one of the five possible matching curves (see Table I for parameter values). Vertical dashed lines show λ values for each model, with curves/dashed lines of matching color corresponding to the same λ value. The maximum value of each curve occurs at the λ value used to generate its samples, as expected. Plot (b) as for plot (a), but using the proportional copying models with parameter b used to specify model [see Table I for (b, c) values].

The existence of isoglosses (spatial linguistic boundaries) suggests that conformity, which creates surface tension at boundaries, may play an important role in language evolution. We have explored this possibility by comparing the spatial distributions of linguistic features created by proportional copying, and by conformity driven copying, to the Survey of English Dialects. We were motivated to seek connections between spatial universality classes in two dimensional physical models (Ising and Voter), and the copying behavior of humans in two dimensional domains. An advantage of comparing models to traditional folk speech is the relative lack of mobility and long range connectivity of speakers, compared to the present day. If speakers do not *physically* diffuse to a great extent, then proportional copying should produce maps of linguistic variants which have similar properties to the spatial distributions generated by the voter model (noise driven interfaces and logarithmic matching functions), whereas conformity-driven copying should produce spatial distributions like those of the Ising model (well defined and relatively smooth interfaces and exponentially decaying matching functions). We have used a spatial language model with copying behavior defined via a learning function which captures how speakers respond to the language state of their community, and is analogous to the activation function of a Hopfield network [37,38]. By choosing this function to be either the identity (proportional copying), or conformity driven, we have been able to generate fictitious language surveys for the English dialect domain which fall into Ising and Voter classes. These maps have then been compared to the Survey of English Dialects [43].

We observed that proportional copying requires speakers to be geographically very isolated (an interaction range of around 100 m), or very small in number, in order to produce the significant spatial variations seen in survey data. This would imply that speakers tend to get their linguistic behavior only from their closest neighbors and immediate family. An alternative explanation is that language communities evolve by proportional copying but with an *effective population* which is much smaller than the true population. This might occur, for example, in a social network dominated by a small number of very influential individuals for whom $\sum_i \omega_{ij} \gg 1$. We do not rule out either of these possibilities,

so we compare our two copying rules based purely on the spatial distributions that they generate. We have shown first that proportional copying tends to generate survey maps with low spatial autocorrelation, as measured by Moran's I . Conformity (interface) driven evolution produces higher I values, consistent with the majority (82%) of survey maps. Second, we constructed statistical models over the space of possible survey maps, derived from theoretical matching curves (logarithmic and exponential) generated by our two activation functions. A likelihood analysis revealed that the majority of maps (80%) were more likely to have been generated by conformity driven evolution, according to our model.

Our work suggests that some form of social conformity is likely to have played a role in the evolution of English folk-speech and, if we accept the SED as a representative dataset with which to investigate language more broadly, then conformity, and surface tension [10], are likely to have played a role in the evolution of other languages.

Code and data availability

All the data and computer code used to generate the results in this paper are available in the publicly accessible GitHub repository Hopfield-SED.

ACKNOWLEDGMENT

The authors are grateful for the Royal Society APEX Award APX180117 which supported this work.

APPENDIX A: CONTINUOUS TIME APPROXIMATION

To derive a continuous time approximation to the spatial model (14) it is useful to write the cell noise as a sum of deterministic and stochastic terms

$$N^{-1}\mathbf{Y}(\mathbf{r}) = \mathbf{v}(\mathbf{r}) + \frac{\boldsymbol{\epsilon}(\mathbf{r})}{\sqrt{N}}, \quad (\text{A1})$$

where the statistical properties of the stochastic term may be understood using the normal approximation to the multinomial distribution [69] (see Appendix B for details). Let \mathbf{O}_v be an orthogonal matrix ($\mathbf{O}_v^T = \mathbf{O}_v^{-1}$) whose last column is \mathbf{v} , and

define $\mathbf{Z} = (Z_1, Z_2, \dots, Z_{q-1}, 0)^T$ where $Z_i \sim \mathcal{N}(0, 1)$, then

$$\boldsymbol{\epsilon} \approx \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v \mathbf{Z}, \tag{A2}$$

where \odot denotes the Hadamard (elementwise) product and $\odot^{\frac{1}{2}}$ is the Hadamard square root. For example, in the binomial case $q = 2$, this yields

$$\boldsymbol{\epsilon} = \sqrt{v_1(1-v_1)} Z_1 \begin{bmatrix} -1 \\ +1 \end{bmatrix}, \tag{A3}$$

where we made use of the fact that $v_1 = 1 - v_2$. Approximating the noise terms from nearest neighbor cells with their mean values, we obtain

$$\delta \mathbf{u}(\mathbf{r}) = \frac{\tau_s}{\tau_m} \left(\underbrace{\mathbf{v}(\mathbf{r}) - \mathbf{u}(\mathbf{r})}_{\text{Nonlinearity}} + \underbrace{\frac{\sigma^2}{2} \nabla^2 \mathbf{v}(\mathbf{r})}_{\text{Diffusion}} + \underbrace{\frac{\boldsymbol{\epsilon}(\mathbf{r})}{\sqrt{N}}}_{\text{Noise}} \right). \tag{A4}$$

If $\mathbf{W} = (W_1, W_2, \dots, W_{q-1}, 0)^T$ is a vector of standard Brownian motions then

$$\boldsymbol{\epsilon} \stackrel{d}{=} \frac{1}{\sqrt{\tau_s}} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v \int_{t-\tau_s}^t d\mathbf{W}_t, \tag{A5}$$

where $\stackrel{d}{=}$ denotes equality in distribution. Defining the continuous time differential form of $\boldsymbol{\epsilon}$ via

$$\boldsymbol{\epsilon} \triangleq \int_0^{\tau_s} d\boldsymbol{\epsilon}_t, \tag{A6}$$

we have

$$d\boldsymbol{\epsilon} = \frac{1}{\sqrt{\tau_s}} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v d\mathbf{W}. \tag{A7}$$

We now suppose there exists a continuous time function $\hat{\mathbf{u}}_t(\mathbf{r})$ such that

$$\delta \mathbf{u}(\mathbf{r}) = \int_{t-\tau_s}^t d\hat{\mathbf{u}}_t(\mathbf{r}). \tag{A8}$$

We now note that $\mathbf{u}(\mathbf{r})$ may be viewed as a piecewise constant function of continuous time which changes at times $\tau_s, 2\tau_s, 3\tau_s, \dots$, then (A4) may be written

$$\int_{t-\tau_s}^t d\hat{\mathbf{u}}_s(\mathbf{r}) = \frac{1}{\tau_m} \left(\mathbf{v}(\mathbf{r}) - \mathbf{u}(\mathbf{r}) + \frac{\sigma^2}{2} \nabla^2 \mathbf{v}(\mathbf{r}) \right) \int_{t-\tau_s}^t ds + \frac{\tau_s}{\tau_m \sqrt{N}} \int_{t-\tau_s}^t d\boldsymbol{\epsilon}_s, \tag{A9}$$

which has differential form

$$d\hat{\mathbf{u}} = \frac{1}{\tau_m} \left(\mathbf{v} - \mathbf{u} + \frac{\sigma^2}{2} \nabla^2 \mathbf{v} \right) dt + \frac{\sqrt{\tau_s}}{\tau_m \sqrt{N}} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v d\mathbf{W}, \tag{A10}$$

where \mathbf{r} dependence has been omitted for brevity. If changes in $\hat{\mathbf{u}}_t(\mathbf{r})$ over the discrete time steps are small, then $\mathbf{u}(\mathbf{r}) \approx \hat{\mathbf{u}}_t(\mathbf{r})$ and we obtain the continuous time approximation (15) given in Sec. II B.

APPENDIX B: NORMAL APPROXIMATION TO MULTINOMIAL

We review the normal approximation to the multinomial distribution (see [69] for more details).

Theorem 1. Let $\mathbf{Y} \sim \text{multinomial}(N, \mathbf{v})$, and define the standardized form

$$\mathbf{Y}^* = \left(\frac{Y_i - N v_i}{\sqrt{N v_i}} \right)_{1 \leq i \leq n}. \tag{B1}$$

Also define the unit vector $\mathbf{u} = (\sqrt{v_1}, \dots, \sqrt{v_n})^T$. Let \mathbf{O}_v be an orthogonal matrix ($\mathbf{O}_v^T = \mathbf{O}_v^{-1}$) whose last column is \mathbf{u} . If Z_1, Z_2, \dots, Z_{n-1} are i.i.d. standard normal variates, then

$$\mathbf{Y}^* \rightsquigarrow \mathbf{O}_v \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_{n-1} \\ 0 \end{bmatrix} \tag{B2}$$

as $N \rightarrow \infty$.

A proof of this theorem is given in Ref. [69]. An intuitive understanding may be obtained by noting that \mathbf{Y}^* lies in the $n - 1$ dimensional hyperplane

$$H_v = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \sqrt{v_i} x_i = 0 \right\}, \tag{B3}$$

so $\mathbf{O}_v^T \mathbf{Y}$ belongs to the hyperplane

$$H = \{ \mathbf{x} \in \mathbb{R}^n : x_n = 0 \}, \tag{B4}$$

which contains the vector $\mathbf{Z} \triangleq (Z_1, \dots, Z_{n-1}, 0)^T$. The matrix \mathbf{O}_v therefore rotates H into H_v (\mathbf{Z} to \mathbf{Y}^*). To practically use this approximation, we require the matrix \mathbf{O}_v , which may be constructed via the Householder transformation. Let \mathbf{w} be a real unit vector, then the Householder matrix

$$P = \mathbf{I} - 2\mathbf{w} \otimes \mathbf{w} \tag{B5}$$

is orthogonal. Defining

$$\mathbf{w} \triangleq \frac{\mathbf{u} - \mathbf{e}_n}{|\mathbf{u} - \mathbf{e}_n|}, \tag{B6}$$

we obtain an orthogonal matrix whose last column is \mathbf{u} . For example, when $n = 2$, we have

$$\mathbf{O}_v = \begin{bmatrix} -\sqrt{v_2} & \sqrt{v_1} \\ \sqrt{v_1} & \sqrt{v_2} \end{bmatrix} \tag{B7}$$

and when $n = 3$, we have

$$\mathbf{O}_v = \begin{bmatrix} \frac{v_2}{1-\sqrt{v_3}} - \sqrt{v_3} & -\frac{\sqrt{v_1 v_2}}{1-\sqrt{v_3}} & \sqrt{v_1} \\ -\frac{\sqrt{v_1 v_2}}{1-\sqrt{v_3}} & 1 - \frac{v_2}{1-\sqrt{v_3}} & \sqrt{v_2} \\ \sqrt{v_1} & \sqrt{v_2} & \sqrt{v_3} \end{bmatrix}. \tag{B8}$$

A more compact statement of theorem 1 may be made by defining \mathbf{E}_n to be the diagonal matrix with a 1 in the first $n - 1$ diagonal entries and 0 in the n th. Letting $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{E}_n)$, then

$$\mathbf{Y} \approx N\mathbf{v} + \sqrt{N} \mathbf{v}^{\odot \frac{1}{2}} \odot \mathbf{O}_v \mathbf{Z} \tag{B9}$$

where \odot denotes the Hadamard (elementwise) product and $\odot^{\frac{1}{2}}$ the Hadamard square root.

APPENDIX C: INTERFACE WIDTH

For conformity driven copying, the steady state shape of the interface, a smoothed step function, solves

$$\frac{\sigma^2}{2}v''(x) = \frac{1}{2}\left[1 + \frac{1}{\beta}\ln\left(\frac{v(x)}{1-v(x)}\right)\right] - v(x), \quad (\text{C1})$$

where $v''(x)$ denotes the lattice second derivative. We assume that $v(x)$ is sufficiently slowly varying so that x may be treated as continuous and (C1) treated as an ordinary differential equation. Without loss of generality we can assume the interface is centered on the origin, where $v(0) = \frac{1}{2}$, which is a fixed point of the right hand side of (C1). As $x \rightarrow \pm\infty$, $v''(x) \rightarrow 0$ and v approaches one of the two other fixed points, which solve

$$\frac{v}{1-v} = \exp(\beta(2v-1)). \quad (\text{C2})$$

We write these solutions, which lie to the left and right of $v = \frac{1}{2}$ as v_-^* and v_+^* . We now define the potential function

$$V(v) \triangleq \frac{(1-2v)^2}{8} - \frac{\ln(2-2v)}{2\beta} - \frac{v\ln\left(\frac{v}{1-v}\right)}{2\beta}, \quad (\text{C3})$$

in terms of which we may write our equilibrium equation (C1)

$$\frac{\sigma^2}{2}v''(x) = -\frac{dV}{dv}. \quad (\text{C4})$$

Noting that $v''(x) = v'(x)\frac{d}{dv}v'(x)$, and integrating (C4) with respect to v , we obtain the conservation law

$$E \triangleq \frac{\sigma^2}{4}(v')^2 + V(v) \quad (\text{C5})$$

where E is a constant, which we may view as a conserved ‘‘energy.’’ To see this, note that if we interpret v, x as position and time variables, then (C1) describes the motion of a particle of mass $\sigma^2/2$ moving in a potential $V(v)$. Noting that $V(\frac{1}{2}) = 0$, then the gradient of the interface at the origin is given by

$$v'(0) = \sqrt{\frac{4E}{\sigma^2}}. \quad (\text{C6})$$

To find E , we note that $\lim_{x \rightarrow \pm\infty} v'(x) = 0$ so

$$E = V(v_+^*) = V(v_-^*) \quad (\text{C7})$$

$$\sim \frac{1}{8} + \frac{\ln 2}{2\beta} \text{ as } \beta \rightarrow \infty. \quad (\text{C8})$$

The interface width is the reciprocal of this gradient.

APPENDIX D: ANNEALED IMPORTANCE SAMPLING

We estimate partition functions by annealed importance sampling [65]. Here we explain how the method is efficiently applied in our case, adapted from the review [70]. We have a pairwise exponential measure

$$\mathbb{P}(\mathbf{s}) = \frac{1}{\mathcal{Z}} \exp[E(\mathbf{s})], \quad (\text{D1})$$

where

$$E(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\theta\mathbf{x}. \quad (\text{D2})$$

We wish to calculate the partition function \mathcal{Z} . We begin with a starting measure $\mathbb{P}_0(\mathbf{s})$, the model with zero interactions, for which \mathcal{Z}_0 is known

$$\mathcal{Z}_0 = 2^L. \quad (\text{D3})$$

We define a sequence of intermediate measures

$$\mathbb{P}_k(\mathbf{x}) = \frac{1}{\mathcal{Z}_k} \exp[E_k(\mathbf{x})], \quad (\text{D4})$$

where $E_k(\mathbf{x}) = kE(\mathbf{x})/K$ and $k \in \{0, 1, \dots, K\}$. Let $T_k(\mathbf{x}; \mathbf{x}')$ be a transition probability which leaves measure k invariant in the sense that

$$\sum_{\mathbf{x}} \mathbb{P}_k(\mathbf{x})T_k(\mathbf{x}'; \mathbf{x}) = \mathbb{P}(\mathbf{x}'). \quad (\text{D5})$$

In other words, \mathbb{P}_k the steady state of the transition matrix T_k . We then generate sequences of states

$$\mathbf{x}_1 \sim \mathbb{P}_0(\mathbf{x}), \quad (\text{D6})$$

$$\mathbf{x}_2 \sim T_1(\mathbf{x}_2; \mathbf{x}_1), \quad (\text{D7})$$

$$\dots \quad (\text{D8})$$

$$\mathbf{x}_K \sim T_{K-1}(\mathbf{x}_K; \mathbf{x}_{K-1}), \quad (\text{D9})$$

and for each sequence, i , out of M , calculate

$$\omega^{(i)} = \prod_{k=1}^K \frac{\exp(E_k(\mathbf{x}_k^{(i)}))}{\exp(E_{k-1}(\mathbf{x}_k^{(i)}))} \quad (\text{D10})$$

$$= \exp\left[\frac{1}{K} \sum_{k=1}^K E(\mathbf{x}_k^{(i)})\right]. \quad (\text{D11})$$

We then have

$$\ln \mathcal{Z} \approx \ln \mathcal{Z}_0 + \ln\left(\frac{1}{M} \sum_{i=1}^M \omega^{(i)}\right). \quad (\text{D12})$$

To see why this method works, suppose that $\mathbf{x}_k \sim \mathbb{P}_{k-1}$ then

$$\mathbb{E}\left[e^{\frac{E(\mathbf{x}_k)}{K}}\right] = \sum_{\mathbf{x}_k} \mathbb{P}_{k-1}(\mathbf{x}_k) \exp\left(\frac{E_k(\mathbf{x}_k) - E_{k-1}(\mathbf{x}_k)}{K}\right) \quad (\text{D13})$$

$$= \sum_{\mathbf{x}_k} \mathbb{P}_{k-1}(\mathbf{x}_k) \frac{\mathcal{Z}_k \mathbb{P}_k(\mathbf{x}_k)}{\mathcal{Z}_{k-1} \mathbb{P}_{k-1}(\mathbf{x}_k)} \quad (\text{D14})$$

$$= \frac{\mathcal{Z}_k}{\mathcal{Z}_{k-1}}, \quad (\text{D15})$$

so $\exp(E(\mathbf{x}_k)/K)$ is an unbiased estimator of $\mathcal{Z}_k/\mathcal{Z}_{k-1}$. If we have an independent sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ then

$$\mathbb{E}\left[\exp\left(\frac{1}{K} \sum_{k=1}^K E(\mathbf{x}_k)\right)\right] = \prod_{k=1}^K \frac{\mathcal{Z}_k}{\mathcal{Z}_{k-1}} \quad (\text{D16})$$

$$= \frac{\mathcal{Z}_K}{\mathcal{Z}_0}. \quad (\text{D17})$$

As shown by Neal [65], even though the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ is not independent, (D17) still holds, so $\omega^{(i)}$ is an unbiased estimator of $\mathcal{Z}_K/\mathcal{Z}_0$.

APPENDIX E: DESCRIPTION OF VARIABLES

1. Introduction

The data used in this study were taken from the SED *Basic materials* [71] rather than from later atlas publications [72,73]. The *Basic materials* presents unmodified transcriptions of question responses rather than defined linguistic variables with discretized variants. Accordingly, here we describe the variables and variants as we have defined them, with references to where in the *Basic materials* the data were taken from. In some instances, especially for lexical variables, data are taken from a single question and the only analysis required is identifying what lexical item(s) each transcription represents; in others, data must be accumulated across multiple questions. In any case, some variants may have to be excluded or merged to define a binary variable as described in part IV of the paper.

For each variable, we give the following information:

- (1) The two variants;
- (2) Reference to where in the SED the data were taken from;
- (3) A linguistic description of the variable and the change which produced it;
- (4) An identification of which of the two variants represents the innovation and which the conservatism;
- (5) Where possible, an approximate dating of the change and so a rough idea of how long the variation had existed at the point the SED speakers acquired the language, the 1880s and 90s (this is more often feasible for lexical and morphological variables, where the written record typically provides more direct evidence than for phonetic and phonological variables); it is assumed for the purpose of this estimate that the first attestation of a form in writing cannot be less than 50 years after its innovation in speech;
- (6) A description of what variants were merged or excluded to define the binary variable used.

Nonlinguists may wish to consult the glossary at the end of this document for an explanation of some of the specialist terminology used. We use the International Phonetic Alphabet (IPA) [74] for transcription throughout. Note that there are differences between the version of the IPA used at the time of publication of the SED (the 1947 chart) and the modern version (the 1999 chart) and we update SED transcriptions to the modern version.

2. Variables

1. Adder lexical item

Variants: (n)adder, hag-

Reference: IV.9.4

Background. This variable describes the lexical item used for the common European viper. The conservative variant is (n)adder, found in the OE period as *nædre* (note that Bosworth and Toller [75] gloss it only as a general term for snake in this period, whilst the OED suggests it had its more specific meaning already in OE). In written sources *hagworm* is known from the late 15th century according to both the OED and MED (*Catholicon Anglicum* c1475) [76]; however,

it is a Norse loanword (ON *hoggormr* “viper”) and so probably dates back to the period of the Danelaw. Accordingly, the change in question is the borrowing of *hag-* from ON, and we should assume this variation has existed for at least 1000 years.

Reduction. As the focus here is on the lexical item, variants of *adder* with and without the metanalytic *n-* and with different reflexes of OE *-d-* were merged; similarly, different formations from *hag-* (*hagworm*, *hagger*, *hag*) were merged. The occasional instances of other lexical items (some clearly in error) were excluded.

2. Anything lexical item

Variants: anything, aught

Reference: V.8.16

Background. This variable concerns the indefinite pronoun used in the frame *Is _left?*, referring to food. Both variants have existed in some form since the OE period (OE *āwiht*, *ānig þing*), along with a variety of other indefinite pronouns (*hwæt*, *ahwæt*, etc.), but it is not clear that both should be considered fully grammaticalized pronouns at this early point. In dating the variation between them, the question we must answer is at what point this grammaticalization process was complete. Mitchell [77] identifies *āwiht* as a fully grammaticalized pronoun in OE, but notes *ānig þing* as a common collocation [78] rather than listing with other pronouns. Bosworth and Toller [79] agree with this implied distinction, in that they give *āwiht* but not *ānig þing* its own entry. On the other hand, they do give examples of *ānig þing* translating Latin *aliquid* without comment (Mæg ænig þing gódes beón of Nazareth a Nazareth potest aliquid boni esse?; [80]), and in an investigation into the syntax of a variety of OE quantifiers, Roehrs and Sapp [81] analyze both *āwiht* and *ānig þing* as being fully grammaticalized heads. Thus it seems reasonable to understand these as already being equivalent pronouns in OE, implying that variation between them has existed for at least 1000 years.

Reduction. The variant *any* was excluded. The more difficult issue with these data concerns the interpretation of [ɔːt] and many similar forms recorded in Devon, with a couple of instances in each of Somerset and Cornwall. The SED interprets these as a phonological variant of *aught*. However, given the carrier sentence, it also seems possible that they represent *ort* “leavings of any description [...] esp. of food” [82] which the EDD does record as occurring in Devon and marginally in Somerset and Cornwall. The problem with the former interpretation is that these are not locations which otherwise exhibit hyperhoticity; the problem with the latter is that it would be expected to be plural and to appear with a quantifier (i.e., *‘Are there any orts left?’ or similar). The syntactic problems with the *ort* interpretation seem hard to overcome, and so the judgment of the editors of the SED has been followed here and [ɔːt] has been merged into *aught*. However, it does seem likely that the lexical item *ort* played some role in the history of this form (whether through analogical change or by *ort* and *aught* being reanalysed as a single lexical item).

3. Fist lexical item

Variants: fist, nieve

Reference: VI.7.4

Background. This variable describes the lexical item used for a clenched hand. The conservative variant is *fist*, found in the OE period as *fyſt* [83]. The OED and MED agree that the earliest written attestation of *nieve* is at the beginning of the fourteenth century (*Havelok* 1300) [84]; however, it is a Norse loanword (ON *hnefi* ‘fist’) and so probably dates back to the period of the Danelaw. Accordingly, the change in question is the borrowing of *nieve* and we should assume the resulting variation has existed for at least 1000 years.

Reduction. The phonological variant of *nieve* with coda /f/ instead of /v/ was merged into *nieve*.

4. Frog lexical item

Variants: *frog*, *paddock*

Reference: IV.9.6

Background. This variable refers to the lexical item used for the set of amphibians referred to in Standard English as *frogs*. The conservative variant is *frog*, which has existed in this meaning since the OE period [85]. The innovative variant *paddock* is derived as a diminutive of *pad* ‘toad’; across the OED and MED, the earliest written attestation is at the beginning of the 14th century (in the compound *padokpipe* c1300, citing Hunt [86]). Thus we can assume the variation has existed for at least 750 years.

Reduction. [θrɒɡz] understood as a phonological variant of *frog* and so merged with it. An additional variant, *jacky(toad)* (apparently derived from earlier *Jacob* ‘frog’ [87]), was excluded on the basis that it is recent, geographically very limited, and entirely embedded in the *frog* domain.

5. Hedgehog lexical item

Variants: *hedgehog*, *urchin*

Reference: IV.5.5

Background. This variable concerns the lexical item used for the European hedgehog, *erinaceus europaeus*. *Urchin* is a loanword, having been borrowed from Norman French *hirchoun* and first attested in English sources around the turn of the 14th century (*South English legendary*, c1300) [88], whilst *hedgehog* is a compound formed within English and first attested in the middle of the 15th century (*Treatise on Fishing*, c1450) [89]. On this (rather limited) basis we might label *hedgehog* the innovation, but in reality both of these were innovations which competed in replacing earlier *igil*, so this is not a particularly useful framing. For this reason, it is not clear that it would be meaningful to give an age for this variable.

Reduction. Compounded and modified variants (*prick(l)(y) urchin*, *pricky black urchin*, *prick(l)y-back(ed) urchin*) were merged into *urchin*; minor variants entirely embedded in the *hedgehog* domain (*hedgепig*, *hedgeboar*, *furzepig*) were excluded.

6. Newt lexical item

Variants: *eft* and related variants, *ask* and related variants

Reference: IV.9.8

Background. Both variants are attested in OE glossing Latin *lacerta* ‘lizard’ [90]. However, *ask* (OE *ādexe*) has cognates elsewhere in Germanic whereas *eft* (OE *efete*) is of unknown origin; additionally, according to the OED, attestations of *ādexe* are found in early OE glosses whereas *efete* is not known until the beginning of the 11th century (OED; [91]). Thus we can take the coining of *eft* to be the

innovation (whether it was a loanword or derived from some other lexical item), and the variation to have existed for at least 1000 years.

Reduction. The raw data contain a great variety of variants. However, many of these are phonological derivations from *eft*, with (*newt*, *mewt*) and without (*ewt*, *eff*, *ebbet*, *abbet*) metanalytic *n-*; these, along with compounded variants (*water-efet*, *wet-efet*, *wet-eff*, *four-legged evet*, *four-legged emmet*), were merged into *eft* on the basis that they imply the earlier use of some form of *eft*. Other variants (*askerd*, *askel*, *asker*) are suffixed forms from *ask*, phonological derivations of these (*askert*, *asgel*, *aster*, *nasgel*) and compounds (*dry-ask*, *water-ask*) and all of these were merged into *ask* by the same logic. Unrelated minor variants (*mancreeper*, *swift*, *water-swift*, *wat-terlizard*, *padgy-pol*, *tiddlywink*, *yellow-belly*) were excluded.

7. Owl lexical item

Variants: *owl*, *howlet*

Reference: IV.7.6

Background. *owl* is an inherited Germanic word that occurs in OE as *ūle* (OED; [92]). The innovation is *howlet*, a loanword from French *hulotte* and first attested in English in the late 15th century (“*Holy berith beris...*” 1475, *Ludus Coventriae* 1475) [93]. Thus the variation can be taken to be at least 450 years old.

Reduction. The key feature by which to distinguish these two variants was taken to be the presence of the second syllable, and so variation in the vowel of the first syllable (short *ullet* versus long *howlet*, etc.), the presence of initial /h/ (*howl* versus *owl*, etc.) and rhoticity in the second syllable (*ullet* versus *ullert*) were ignored. Compounded variants were merged with their respective un-compounded variants (*Jenny-owl* and *Meg-owl* with *owl*, *Jenny-howlet* and *Polly-howlet* with *howlet*).

8. Pour lexical item

Variants: *pour*, *teem*

Reference: V.8.8

Background. This variable refers to the lexical item used for decanting tea from a teapot into a cup. *Teem* is first attested at the beginning of the 15th century (*Cursor Mundi* 1400) [94] but as an Old Norse borrowing (cf., ON *tæma* ‘empty’) we can assume it dates back much earlier. Thus *pour* is probably the innovation: according to the OED and MED it is probably a loanword from Middle French *purier* and is first attested in the first half of the 14th century (*Amis and Amiloun* c1330) [95]). Thus we can take the variation to have existed for at least 600 years.

Reduction. Assorted minor variants were excluded from consideration: *birle*, *chuck*, *emp*, *ent*, *hale*, *hell*, *heave*, *laden*, *lade*, and *shut*.

9. Snail lexical item

Variants: *snail*, forms including *-dod-*

Reference: IV.9.3

Background. This variable refers to the lexical item given in response to the question: “What are those slow, slimy things that carry their houses about with them; they come out after rain?”, intended to elicit the name for terrestrial molluscs with spiral shells large enough to retract into (standard English *snail*). The conservative variant is *snail*, which is an inherited Germanic term known in English from the OE period. The

innovation is the use of terms including the element *-dod-*; since the etymology of *-dod-* is unknown, the nature of this innovation (as borrowing versus derivation from some pre-existing element) is uncertain. These are known from the EMoE period according to the OED (*dodman* in John Bale's *King Johan* 1528; *hodmandod* and *dodman* in Bacon 1626 [96]), so we can assume the variation has existed for at least 400 years.

Reduction. All compounded variants containing *-dod-* (*hodmedod*, *hoddy-doddy*, *dodman*, etc.) were merged as a single variant. Since the etymology of *-dod-* is not known, it is not clear what the precise sequence of derivation was here: **dod* itself might have been a lexical item meaning snail which spread to this region, making this the innovation and all the compounded forms later derivations; or **dod* might have had some other meaning (the OED suggests a connection with *dod* 'rounded summit'), implying that one of the compounds was the original innovation and the others analogical formations based on it. Either way, however, it seems reasonable to regard the use of terms with *-dod-* as an innovation across this region.

10. Upstairs lexical item

Variants: *upover*, *upstairs*

Reference: V.2.5

Background. This variable concerns the lexical item used to describe a room in an upper floor. The innovative variant *upover* is not listed in the OED; in the EDD, examples are cited from Devon, but only from 1877 [97]. Thus this innovation can be assumed to be very recent.

Reduction. Minor variants, quite possibly reflecting failures to elicit the relevant term, were excluded: *up a height*, *up above*, *up top*, and *up aloft*.

11. Vinegar lexical item

Variants: *alegar*, *vinegar*

Reference: V.7.19

Background. This variable concerns the lexical item used for acetic acid solution used in cooking, elicited as "that sour liquid you pickle red cabbage in." Of these two forms, *vinegar* was a loanword from OF *vinaigre*, first attested in this meaning in the first half of the 14th century (*South English Legendary* 1325, *Shoreham Poems* 1350) [98]. *Alegar* appears to be an analogical formation *ale+eager* on the basis of *vinegar* attested from the end of the 14th century (*Form of Curry* 1399, *Inventories of St. Leonard's Priory* c1422) [99]. However, since *vinegar* referred specifically to wine vinegar and *alegar* to malt vinegar, the relevant innovation is not the coining of either word but the shift of *alegar* to overlap in meaning with *vinegar* so that the two could be considered variants of a single variable. None of the reference materials consulted here record this meaning, suggesting that this innovation was relatively recent (although in many written contexts it would be very hard to identify the change).

12. Wrist lexical item

Variants: *wrist*, *shackle*

Reference: VI.6.9

Background. This variable concerns the lexical item used for the end of the arm before the hand. Both words are native Germanic, found already in OE (*wrist*, *sceacel*) and with cognates in other Germanic languages. However, *shackle* in the meaning 'wrist' is a shortening of the compound *shacklebone*

'wrist' and so it is the formation of this compound that is the relevant innovation; the first instance recorded in the OED is from the third quarter of the 16th century (*Register of the privy council of Scotland* 1571), so we can assume this variation is at least 350 years old.

Reduction. Compounded variants with *-wrist* (*armwrist*, *handwrist*) were merged into *wrist*.

13. Yeast lexical item

Variants: *yeast*, *barm*

Reference: V.6.2

Background. This variable concerns the lexical item used for the substance added to bread dough to raise it. The two lexical items concerned, *yeast* and *barm*, are both native Germanic words which existed in OE (*gist*, *beorma*). In modern standard English, there may be a semantic distinction between the two words whereby *barm* is used to refer to the foam removed from the top of fermenting malt liquors whilst *yeast* refers to the fungus that causes fermentation; historically, however, both terms had the former meaning (OED; [100,101]). Thus we should understand the variation between these two variants as having existed for at least 1000 years.

14. Participial adjective from burn

Variants: *burnt*, *burned*

Reference: V.6.7

Background. This is a morpholexical variable: it is part of a wider pattern of morphological variation between (regular) /d/ and (irregular) /t/ suffixes for forming the preterite, but the choice is lexically controlled for the individual speaker and the variation is independent across lexical items. The verb *burn* is descended from two OE verbs, strong class III *beornan* and weak class 1 *baernan*, which merged during the ME period, along with admixture with parts of the two corresponding ON verbs (strong intransitive *brenna* and weak transitive *brenna*) (OED; [102]); the MoE past tense forms are clearly only related to the weak formations. The innovative /t/ variant does not seem to occur in OE (it is not mentioned in Bosworth and Toller [103] and there are no occurrences in the Old English Web Corpus [104]). The first ME occurrences according to the MED [105] and LAEME [106] are in the late 13th and early 14th centuries (Cambridge Trinity B.14.39 (=LAEME text #246) after 1253, *Genesis & Exodus* 1325), so we can assume the variation is at least 750 years old.

Reduction Excluded responses with unrelated verbs (*kizzened*, *swinged*).

15. Past participle of earn

Variants: *earnt*, *earned*

Reference: VIII.1.26

Background. This is a morpholexical variable: it is part of a wider pattern of morphological variation between (regular) /d/ and (irregular) /t/ suffixes for forming the preterite, but the choice is lexically controlled for the individual speaker and the variation is independent across lexical items. The verb *earn* is descended from a weak class 2 OE verb *earnian*; as such, the regular variant (*earned*) is conservative and innovative *earnt* must be by analogy with another verb. The irregular variant appears first in the written record in the 18th century (early attestations include: Anon. 1730 [107]; Smith 1737 [108]; Ward 1758 [109]; Nugent's 1763 translation of Rousseau [110]), so we can assume the variation is at least 300 years old.

16. Past participle of *get*: presence of *-en*

Variants: $-\emptyset$, *-en*

Reference: IX.6.4

Background. This is part of broader variation in the morphology of English past participles, but the variation occurs at the level of individual lexical items and so should be classed as morpholexical. The OED notes that in varieties with both variants *got* and *gotten* there is often a semantic distinction where *have gotten* refers to the process of obtaining something whilst *have got* refers to simple possession (as might be expected for an ongoing grammaticalization process). The framing sentence used in the SED is: *You say to a friend: Shall I give you one of these pups? But he answers: No thanks, we _one.* This perhaps leaves open space for either interpretation, but simple possession seems more likely. The verb *get* is primarily descended from the ON class V strong verb *geta*, perhaps with some influence from its OE cognate *gietan* (OED). The $-\emptyset$ variant is the innovation. Forms without the final /n/ (i.e., *-e*) are found as a result of general final /n/ loss in ME as early as the late 14th century ('*gote*' in *Wycliffe Bible* 1382) but since the MoE vowel is short these are unlikely to be precursors of the $-\emptyset$ form. The MED lists *g(h)et* and *gat* as possible forms of the past participle, but the only examples given are one instance of *gat* at the beginning of the 15th century [*Cleanness* (Nero A.10) c1400] which context renders ambiguous between a past participle and a simple preterite, and one of *geth* in the mid-15th century (*Paston letters* 3.2 1454) [111]. Thus the earliest we can confidently date the innovation to is the early 15th century, rendering it at least 600 years old.

Reduction. Variation in the stem vowel was ignored, so that *getten*, *gitten* and *gotten* were merged as one variant and (*a*)*got*, *gat*, and *got* as the other.

17. Preterite of *grow*

Variants: *growed*, *grew*

Reference: IX.3.9

Background. This is a morpholexical variable: it is part of larger patterns of morphological variation between weak and strong preterites, but the choice of variant is lexical and not correlated across different verbs and locations. The strong form *grew* is the conservative variant and attested from the OE period, whereas the weak form *growed* is an analogical formation first attested in ME from the latter half of the 14th century (*William of Pallern* 1375, *Wycliffe Bible* 1382) according to citations in the MED [112]; accordingly we can assume the variation is at least 650 years old.

Reduction. The SED records an occasional third variant, *did grow*, in localities such as So4 and Ha6. However, these are likely to reflect examples of habitual *do* and not truly a variant of the simple preterite; accordingly, these were excluded.

18. Backformation of sg. *pea*

Variants: sg. *pea* pl *peas*, sg. *pease* pl *pease*

Reference: V.7.13

Background. This morpholexical variable concerns variation in the paradigm of *pea(se)*. The noun is originally weak, with plural in *-n* (OE *pise* : *pisan*, ME *pese* : *pesen*). When weak plurals were lost in ME, it gained a strong plural in a handful of dialects (e.g., *peses* in *Piers Plowman* C 9.307, c1400) but in most varieties the result were identical singular

and plural forms *pease* : *pease*. The final /z/ of the stem was later reinterpreted as plural *-s* and so a singular form *pea* was backformed from it. The earliest attestations of this form in the OED are from the mid- to late 17th century [113–115], so we can assume the variation is at least 400 years old.

19. Worse: formation of *worser*

Variants: *worse*, *worser*

Reference: VI.12.3

Background. This morpholexical variable refers to the formation of the comparative of *bad* as *worse* versus *worser*. This word probably did originally have a distinct comparative suffix (Gothic has *wairsiza*, on the basis of which Magnússon reconstructs PG **werz-izan-* [116]) but by the OE period this is no longer synchronically recognizable (OE *wiersa*) making *worse* the conservatism. The suffix *-er* is then added to this by analogy with regular comparatives in some varieties in the late ME/EMoE period. The earliest attestations in the OED are from the end of the 15th century and beginning of the 16th (including *De proprietatibus rerum* 1495, *Mirk* 1508 [117]) so we can assume the variation dates back at least 550 years.

Reduction. The unrelated lexical variant *waur* was excluded.

20. Worse: lexical item

Variants: *worse*, *waur*

Reference: VI.12.3

Background. This variable refers to the lexical item used suppletively as the comparative of *bad*: either native *worse(r)* (<OE *wiersa*) or the borrowing *waur* (< ON *verri*). The earliest attestation of the loanword in citations in the OED and MED is in the late 12th century (*Ormulum* 1175) [118] implying that the variation has existed for at least 850 years, but, as for any ON loanword, identifying a terminus post quem in this way is likely to give an underestimate of the age of the variation: we can assume that this was borrowed during the period of the Danelaw, suggesting an age of at least 1000 years.

Reduction. The variant *worser* was merged with *worse*, since *worser* is transparently a later derivation from *worse*.

21. Possessive pronouns

Variants: *-s*, *-(e)n*

Reference: IX.8.5

Background. This is a morphological variable referring to the formation of the possessive personal pronouns. The histories of the individual person-number forms should initially be considered separately.

For the 3sg. feminine, both variants are ME analogical constructions based on the 3sg. feminine personal pronoun *hire* and the possessive pronouns *mīn*/*pīn* in the case of *hiren*, and gen.sg. *-es* in the case of *hires*. According to citations in the MED [119], *hiren* is attested as early as the first half of the 13th century (*Ancrene Riwe* c1230) whereas *hires* is not attested until the fourth quarter of the 14th century (*Wycliffe Bible* 1382), suggesting that we should see *hires* as the innovation; this is consistent with the fact that the expansion of gen.sg. *-es* was not complete until the end of the ME period [120].

For the 3sg. masculine, the *-s* variant is the original form and found regularly since the OE period [121]. The *-n* variant *hisen*, like *hiren*, is an analogical formation based on the 1sg./2sg. possessive pronouns, but is not attested until the

mid-15th century according to the MED (*Laud Troy Book* c1425, *Letters pertaining to the Guilds of Coventry* 1440) [122].

For the 1pl, both variants are analogical formations and instances of both are found from the late 14th century (*ourn* in *Wycliffe Bible* 1382, *oures* in the *Pardoner's Tale* 1390) [123].

For the 3pl, as with the 3sg. feminine and the 1pl, both variants are analogical formations. At least one instance of the *-s* variant is found as early as the late 12th century (*Ormulum* c1175) but it starts to appear regularly only from the late 14th (*Wycliffe Bible* 1382, *Cursor Mundi* 1400) [124]. The *-n* variant appears to be substantially later: the MED notes just one example, in the mid-15th century (*Treatise on the Ten Commandments* c1425) and suggests this may be a secondary analogical formation based on the 3sg. feminine instead of being by analogy with the 1sg./2sg. [125].

We can see that the two variants have somewhat different histories and exact dating for the different persons/numbers. However, the innovation of these systems, in which the possessive pronouns are all formed with *-s* or are all formed with *-n*, can be given termini post quem by looking at the latest forms to appear: the mid-15th century for the *-n* variant, the late 14th century for the *-s* variant. Thus the variation is at least 650 years old.

Reduction. The 3sg. feminine, 3sg. masculine, 1pl and 3pl are treated together, so that *hers*, *his*, *yours*, and *theirs* are merged as *-s* and *hern*, *hism*, *ourn* and *theirn* are merged as *-n*. Double marked variants (*hersn*, *ourns* etc.) could equally be derived from earlier *-s* or *-n* variants, and so were excluded. Zero marked variants (*her*, *our* etc.) probably reflect a misunderstanding of the question and so were also excluded.

1sg. *mine* and 2sg. *thine* do not participate in this system (they always have *-n*), and so are not included. 2sg. *yours/yourn* is not recorded in many localities where *thine* is still used, and so is not included; additionally, as the use of the historical plural in the singular is more recent than the innovation of this variable and (in recent decades) reflects spread from Standard English, it is not clear that we would expect it to be part of the same system. 2pl *yours/yourn* has a substantially different distribution to the other person/number combinations, presumably again reflecting this interaction with Standard English and with the 2sg., and so was not included.

22. Verbal 3sg. *-s*

Variants: *-s*, \emptyset

Reference: VI.5.5 (*speaks*), VI.13.3 (*aches*, *hurts*), VI.14.2 (*suits*), VI.14.14 (*wears*), VIII.1.9 (*looks*, *favors*, *resembles*), VIII.6.2 (*begins*, *breaks*, *closes*, *comes*, *finishes*, *leaves*, *opens*, *shuts*, *starts*), IX.3.6 (*makes*)

Background. This morphological variable concerns the form of the verb used with a 3sg. pronoun subject. There has been variation in subject-verb agreement at least since the OE period, with Northumbrian texts such as the 10th century Lindisfarne Gospels showing variable *-es* for all persons/numbers alongside more conservative forms (see, e.g., Refs. [126]). This *-es* ending spread from northern to southern English varieties throughout the ME period and became restricted to the 3sg., rising in frequency dramatically in London English in the late 16th century [127,128]. The zero ending may be the result of analogical levelling across the paradigm or may have

its origin in subjunctive zero endings. Either way, it existed at a very low frequency in many EMoE varieties, but became particularly established in East Anglia from the 16th century; it was also found particularly in parts of the south west of England, perhaps as a result of the changes involving positive declarative *do* (for these points and further, see Wright [129]).

As can be seen from this brief account, it is not straightforward to identify a conservative and an innovative variant here. Both endings have existed for a very long period of time, but their functions and the roles they play in the larger inflectional system have shifted. We first see them occurring in systems that look broadly like the MoE systems (with levelled \emptyset in all person/number combinations on the one hand, or with *-s* distinguishing the 3sg. from all other cells in the other) in the south of England at roughly the same period of EMoE. Accordingly, it seems reasonable to think of this variation as around 500 years old.

Reduction. Instances of habitual *do*+verb were excluded from consideration.

23. 1sg. present of *be*: *levelling to be*

Variants: levelled to *be*, not levelled to *be*

Reference: IX.7.1

Background. The verb *to be* in Standard English differs from all other verbs in showing a pattern of subject-verb agreement that distinguishes more than just 3sg. versus other. In traditional dialects, this system is simplified in a large variety of different ways. In this variable, we look just at the levelling of 1sg. (standard English *am*) to *be*, but for most speakers that reflects a system with levelling to *be* in all person/number combinations. In OE there were two verbs meaning 'be': *wesan* and *bēon*, of which *wesan* was unmarked whilst *bēon* was typically used for the gnomic present, the future, or the iterative present/future, with many exceptions [130]. It seems likely that MoE dialectal systems with levelling to *be* date back to the collapse of the *wesan:bēon* system, rather than being a later development: the innovation, under this understanding, is levelling of *wesan* forms to *bēon* as the semantic distinction between them was lost.

LAEME offers some evidence for *be* forms used in the 1sg. with future meaning in the Midlands [131], but has no map for 1sg. *be* forms used in other contexts. There is no evidence of this form in southern ME in LALME [132]; however, the MED lists examples in the indicative from the middle of the 15th century (*King Ponthus* 1450, *Pilgrimage of the Life of Man* 1500) [133]. By contrast, LAEME, LALME and the MED offer copious evidence for *be-* forms in the 2sg., 3sg., and pl (indeed, *be-* forms are universal in the pl in southern ME), suggesting that spread to the 1sg. was the last stage of this levelling process. Together with the fact that a system with complete *be-*levelling must have existed by the end of the 16th century since it was part of the input to Caribbean Englishes, this suggests that we can date this innovation to some time in the 15th century, making the variation at least 550 years old.

Reduction. The variants *be* and *bin* were merged as showing *be-*levelling; the variants *am*, *are* and *is* were merged as not showing *be-*levelling.

24. 1sg. present of *be*: *levelling to is*

Variants: levelled to *is*, not levelled to *is*

Reference: IX.7.1

Background. The verb *to be* in standard English differs from all other verbs in showing a pattern of subject-verb agreement that distinguishes more than 3sg. versus other. In traditional dialects, this system was simplified in a large variety of different ways. In this variable, we look just at the levelling of 1sg. (standard English *am*) to *is*. Historically, this variant might have been related to the pattern known as the “North subject rule” by which present tense verbs took the 3sg. *-s* form in all contexts except where they were directly adjacent to a personal pronoun; the NSR could have generated 1sg. *is* when the subject was not adjacent to the verb which might later have been extended to other contexts, and was associated with a similar spatial region to that we see in the SED for this variable. However, this has not been investigated in detail. There are no tokens of 1sg. *is* in LALME [134], but LALME only has data for southern England for this variable. LAEME does not map this variable specifically; exploring the tag dictionary we find one text with relevant examples, 1sg. ⟨*es*⟩ in hand C of the 14th century *Cursor Mundi* [135], but these reflect just two tokens in this long text which otherwise uses ⟨*am*⟩. In a study of early evidence for the NSR, de Haas finds that NSR with full verbs dates from as early as the 10th century [136]; but this study excluded *to be*, meaning it offers no direct evidence for this variable [137]. We have not been able to identify any occurrences in the Parsed Corpus of Early English Correspondence [138], which covers the period 1410–1681. Overall, then, all we can say about the age of this variant is that it likely has its origins in NSR which is of OE or EME age and that it may have existed in some form since the ME period, but that we do not have clear enough evidence to offer a specific date.

Reduction. The variants *am*, *are*, *be*, and *bin* were merged as not showing *is*-levelling.

25. It is contraction

Variants: *'tis*, *it's*

Reference: V.7.3

Background. This morphological variable concerns the contracted form of the 3sg. inanimate pronoun, *it*, plus the 3sg. of *to be*, *is*, in phrase-internal position. The verb *to be* has exhibited contractions with various pronouns and negative adverbs since the OE period, but contractions of *(h)it+is* in particular seem to go back to the late ME period. The OED lists examples of *tis* as early as the late 13th century (*Ancrene Riwle* 1289) but without syncope (i.e., *hit tis*); in both the OED and MED, the first cited example with syncope of the vowel of *is* is from the latter half of the 15th century (*Mankind* c1475) [139]. There are no examples with syncope of the vowel of *it* in the MED, and the earliest cited instance of *it's* in the OED is in the mid-16th century [140], suggesting a point of innovation at some time in the 16th century. These dates suggest that *tis* was the conservatism, well-established by the time that *it's* was innovated; the relative trajectory of the two variants in printed materials in Google Books [141] supports this, cf., Fig. 19. Thus we can infer that this variation has existed for around 350 years.

Reduction. Responses with no vowel (i.e., [ts]) were excluded, as they are ambiguous between *it's* and *tis*.

26. Coda /l/ velarization

Variants: clear [l], dark or vocalized [ɫʊ]

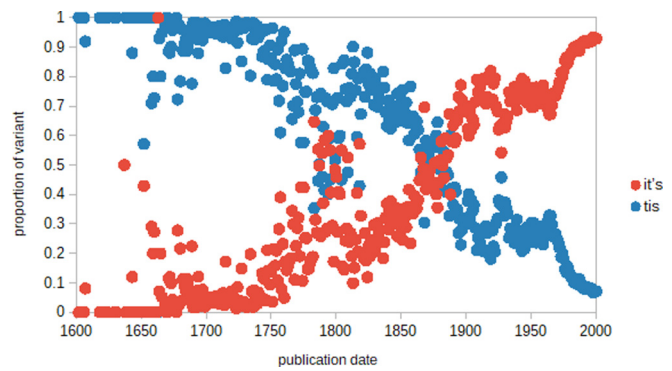


FIG. 19. Relative rate of *'tis* and *it's* in the Google Books corpus by date.

Reference: IV.7.6 (*owl*), V.9.7 (*shelf*), VII.3.7 (*fall*), VII.6.10 (*dull*)

Background. This variable concerns velarization (and potentially subsequent vocalization) of /l/ in coda position. There is evidence for this sound change throughout the history of English (and, indeed, its reconstructed prehistory, if vocalization of proto-Indo-European syllabic liquids is taken into account). Sporadic instances from the EME period give MoE forms like *which*, *such* and *as*; systematic occurrence in the frame [V+back]_[C+labial, C+back] starting in the north of England from the 15th century onwards gives MoE forms like *yolk*, *half*, and *folk* [142]; systematic occurrence after back vowels regardless of following context gives Modern Scots forms like *a' 'all*, *pou 'pull*, and *fou 'full* [143]. However, these earlier instances of the sound change are excluded from consideration here, and only the most recent occurrence is examined, which applies to all coda /l/ regardless of preceding vowel (and, although these are not included here, also syllabic /l/) and is associated with the south-east of England. Since this latest sound change rarely affects the orthography, it is difficult to date from written sources.

Reduction. Vocalized realizations, given that they are universally back, must have proceeded via dark [ɫ], and so are merged with it.

27. Coda /st/ simplification

Variants: [s], [st]

Reference: VI.6.9 (*wrist*), VI.7.4 (*fist*), VII.6.6 (*frost*)

Background. This variable concerns the simplification of coda /st/ clusters to [s]. The existence of this process as a fast speech process, but not a regular sound change, is a universal of MoE varieties and can be found throughout much of the history of English. For example, looking at the superlative suffix *-est*, LALME shows just six points with simplification ((-es), (-ys)) and these are scattered evenly across the map, suggesting spelling errors or sporadic sound change rather than a regular sound change [144]. However, the SED offers evidence for a more consistent regular sound change in certain regions, in particular Devon, east Cornwall and West Somerset. It is hard to date this later change specifically.

28. FACE vowel diphthongization

Variants: fronting diphthong, other

Reference: V.1.5 (*gable*), VI.13.3 (*aches*), IX.3.6 (*make*, *makes*, *made*)

Background. This is a phonetic variable, referring to the MoE realization of ME /a:/. Changes affecting ME /a:/ are extremely complex, with a great diversity of different reflexes at the time of the SED. Anderson [145] identifies four major groups, reflecting four sound changes:

(a) Rising diphthongs, reflecting the development of a stressed, high front onset;

(b) Centring diphthongs, reflecting the development of a schwa offglide;

(c) Long monophthongs, reflecting either no changes beyond the raising that affected all ME /a:/, or only further raising;

(d) Upgliding diphthongs, reflecting the development of a high front second element.

Of these, Anderson reasons on the basis of information from Ellis that (d) is a recent development, spreading rapidly from London only in the 19th century (Anderson [146] citing Ellis [147]), during which time other variants have been recessive. Accepting this argument, we examine here the distribution of this latest change.

Reduction. All of the (d) variants [aɪ, æɪ, eɪ, ei, ei, ei] were merged as showing the change in question. All of the (a) [iɑ, iɑ, iə, jaɪ, jeɪ, jeɪ], (b) [ea, eə, eə], and (c) [ɛ, ɛɪ, e, eɪ, i, i] variants were merged as “other.”

29. *h-dropping*

Variants: [h], [∅]

Reference: IV.5.10 (*hare*), IV.10.9 (*holly*), V.1.1 (*houses*), VI.9.1 (*hip, haunch, huck*), VI.10.9 (*height*), VII.6.6 (*hoarfrost*)

Background. The variable refers to the nonrealization of etymological onset /h/. There has been at least sporadic loss of onset /h/ in English since the OE period, but this became a more established part of the phonology of many varieties during the ME period, partly as a result of Romance contact [148]. This history is relatively complex and the evidence equivocal, with different behaviors of native words and loanwords, prescriptive pressures and sociolinguistic effects, and influence from the orthography; however, this is not important for our purposes.

30. *happy lowering*

Variants: lowered or lax vowel, high vowel

Reference: VIII.1.16 (*ready*)

Background. This is a phonetic variable, dealing with realizations of the unstressed, word-final front vowel labeled as the ‘happy’ vowel in Wells’ lexical sets [149]. We see two sound changes reflected in the SED data: one which lowers the vowel to [e] or [ə], and one which tenses it to [i]. Since the distributions of these reflexes do not overlap, they are both included here as separate variables.

Reduction. The high variants [ɪ i] were merged as not showing lowering, the mid variants [e ə] were merged as showing lowering.

31. *happy tensing*

Variants: tensed vowel, lax or lowered vowel

Reference: VIII.1.16 (*ready*)

Background. See happy lowering (30).

Reduction. The lax and lowered variants [ɪ e ə] were merged as not showing tensing.

32. *ME /o:/: EMoE fronting*

Variants: fronted [ɣ, ɣ:], other

Reference: V.1.2 (*roof*), V.2.4 (*rooms*), V.2.14 (*broom*), VI.5.6 (*tooth*)

Background. The SED shows a huge diversity of reflexes of ME /o:/, reflecting a series of overlapping sound changes. We examine three sets of reflexes reflecting three changes whose outputs are phonetically distinct and whose spatial distributions are relatively well separated: the fronting of ME /o:/ to [ø:] (or similar) in the north-west of England, giving MoE reflexes with front falling diphthongs ([ɪə] etc.) [150]; the shortening of long /u:/ to [ʊ] [151]; and the fronting of raised /u:/ to [ɣ:] [152]. The last of these is a phonetic variable; the other two are phonological, in the sense that they result in mergers with other phonemic vowels. The first had happened already in the ME period, whereas the latter two are later. See ME /o:/: shortening (62) and ME /o:/: ME fronting (63) for the other two changes.

Reduction. Front rounded monophthongs [ɣ, ɣ:] were merged as showing the change, all other variants [ɛʊ, aʊ, ʊʊ, ʊʊ, ʊʊ, ʊə, ɪə, ɪə, ɪʊ, ɪʊ, jʊ, jʊ, uɪ, ʊuɪ, ʊu, ʊi, ɪu, ɔuɪ, ɪ, ʊ, ʌ, ʌ:, ɣ] were merged as not showing the change.

33. *ME /u:/: MOUTH monophthongization*

Variants: MOUTH monophthongization, no MOUTH monophthongization

Reference: IV.5.2 (*mouse*), V.1.1(a) (*houses*), VI.14.14 (*trousers*), VII.1.16 (*thousand*)

Background. See ME /u:/: Great Vowel Shift (64).

Reduction. Long low or mid monophthongs [ɑ, aɪ, æ, æɪ, ɛ:] and low or mid front vowels with a schwa or low offglide [a:ə, æ:a, æ:ə, æa, æə, ɛa, ɛə, ea] were merged as showing the change; all other reflexes [ɑɪ, aɪ, əu, ɔuɪ, əʊ, əʊ, ʌʊ, ɛu, ɛʊ, ɛu, ɛuɪ, ɛw, ɛuɪ, ɛʊ, ʊuɪ, uɪ, u, aʊ, au, auɪ, aʊ, ɑu, ɑʊ, æu, æʊ, æ:ʊ, æɣ, ɔɣ] were merged as not showing the change.

34. *Onset /f/ voicing*

Variants: [f], [v]

Reference: IV.9.6 (*frogs*), IV.10.11 (*furze*), V.3.1 (*fire*)

Background. This variable concerns the realization of etymological /f/ as [v] in initial position in native Germanic words. This is part of a larger set of changes by which all initial nonback fricatives could be voiced: /f θ s ʃ/ > [v ð z ʒ]. The chronology and evidence is somewhat different for the different fricatives, however: /ʃ/ > [ʒ] is much more inconsistent than the others and should probably be assumed to be a later, separate change; /f/ > [v] is best evidenced in medieval sources (but possibly only because the orthography had ways of representing [v], viz. < u v w >) and shows a relatively high rate of application in potential items in the SED (66.3%); /s/ > [z] is less well evidenced in medieval sources, but this is expected given that EME orthography did not have a widely accepted orthome for /z/, and it shows a similarly high consistency in the SED (70.3%); /θ/ > [ð] is so poorly evidenced in medieval sources that it is hard to meaningfully trace its distribution, but it is the most consistent in the SED (81.1%) [153]. The first direct spelling evidence for this change is an instance of < u > for /f/ in a document from around 950 [154]. The orthographic evidence then mounts through the EME period; the text which shows the change most consistently is a Kentish text from the first half of the 14th century (*Ayenbite of Inwit* 1340) [155]. There is some disagreement about whether the changes at different places of articulation

should be seen as separate sound changes or whether it might be possible to unify them as a single change (the latter is argued for by Fisiak [156]); however, in either case, it seems clear that later retreat of these isoglosses has proceeded somewhat differently for the different phonemes involved, justifying our treating /f/ > [v] and /θ/ > [ð] here as separate changes.

There are broadly two possible positions on the dating of this change. Either the orthographic evidence is taken as offering us evidence for timing the of the change (at least indirectly and at a delay), implying that it took place somewhere around the EME period; or the timing of the orthographic changes are seen as entirely unrelated to the timing of the sound change, in which case the sound change might have happened much earlier in the OE period. The former “traditional” view is put forward in Brunner [157], Berndt [158] and Pinsker [159], among others. The latter view is argued tentatively by Fisiak [156], more directly by Bennet [160] and Lass [161]. This latter view has the advantage that it allows the sound change to be identified as the same sound change that voiced initial fricatives in varieties of Dutch and Low German. We accept this view, implying an extremely early date: that this variation existed in the speech community since before the migration of Germanic speakers to the British Isles, and thus the data here reflect a geospatial distribution that has been evolving since the migration period.

35. Onset /θ/ voicing

Variants: [θ], [ð]

Reference: V.8.16 (*anything*), V.10.2 (*thread*), V.10.9 (*thimble*), VI.6.3 (*throat*), VII.6.15 (*thawing*), VIII.1.3 (*three*), VIII.7.7 (*throwing*)

Background. See onset /f/ voicing (34).

Reduction. Stopped and fronted variants were merged with dental fricatives according to their voicing (i.e., [f t t̥ t̥^h θ] were merged as [θ] and [v d ɖ ð] as [ð]). This is on the basis of the assumption that the change in voicing preceded other changes affecting this phoneme.

36. Postvocalic /t/ glottalization

Variants: [t], [ʔ]

Reference: V.8.7 (*kettle*), VI.5.11 (*eat, ate, eaten*), VII.3.7 (*autumn*), VIII.1.4 (*daughter*), VIII.1.11 (*brought*), VIII.1.12 (*aunt*), IX.3.8 (*caught*), IX.6.4 (*got, gotten*)

Background. This variable refers to the realization of postvocalic /t/ as a glottal stop [ʔ]. This sound change is characteristic of British English much more than colonial varieties, suggesting a late date of innovation. However, since there are no explicit contemporary commentaries before the second half of the 19th century, dating it specifically is difficult [162]. Here, we take it that there is no reason to date it earlier than the beginning of the 19th century.

Reduction. Glottalized variants were treated together, whether or not they were debuccalized (i.e., [ʔ] and [tʔ] are merged as [ʔ]); voicing and aspiration are ignored, meaning that [t, t^h, ɖ] are merged as [t].

37. Postvocalic /t/ voicing

Variants: [t], [d]

Reference: V.8.7 (*kettle*), VI.5.11 (*eat, ate, eaten*), VII.3.7 (*autumn*), VIII.1.4 (*daughter*), VIII.1.11 (*brought*), IX.3.8 (*caught*), IX.6.4 (*got, gotten*)

Background. This variable refers to the realization of postvocalic /t/ as a voiced stop [d]. This sound change is found in North American English varieties as well as British English varieties, suggesting an early date of innovation.

Reduction. All voiceless variants [ʔ, t̥, t̥^h] are merged as not showing the change.

38. Prevocalic /r/ backing

Variants: uvular [ʁ], coronal [r r̥ r̥̄]

Reference: II.9.1 (*grass*), IV.5.8 (*squirrel*), IV.11.1 (*blackberries, brambles*), V.5.3 (*cream*), V.3.11 (*draught*), VII.4.8 (*Christmas*), IX.3.9 (*grow, grew, growed*)

Background. This phonetic variable refers to the place of articulation of (prevocalic) /r/ as uvular or coronal. Following Minkova [163], we assume that the historically prior realization was an alveolar or dental trill [r] and so the innovation here is the backing of this phoneme to a uvular trill or approximant [ʁ ʁ̄]. The earliest written reference to this sound change dates from 1724 [164], cf., [165], so we assume that this variable has existed for at least 200 years.

Reduction. All of the coronal realizations [r r̥ r̥̄] were merged as a single variant.

39. Prevocalic /r/ retroflexion

Variants: retroflex [ɻ], dental/alveolar [r̄ r̄̄]

Reference: II.9.1 (*grass*), IV.5.8 (*squirrel*), IV.11.1 (*blackberries, brambles*), V.5.3 (*cream*), V.3.11 (*draught*), VII.4.8 (*Christmas*), IX.3.9 (*grow, grew, growed*)

Background. This phonetic variable refers to the realization of prenasal /r/ as retroflex [ɻ] versus dental/alveolar [r̄ r̄̄]. It is generally agreed that the earliest realization of this sound was a dental/alveolar trill (although see Minkova [163] for references to the argument that the uvular variant is historically prior). Tristram [166] argues that the next stage of the sequence of changes was a shift to retroflex place of articulation, that this happened already in West Saxon, and spread during the OE period to the limit of the Danelaw. However, here we instead accept the more parsimonious account that lenition from trill to approximant preceded changes in place [167,168], citing [169,170], and so the retroflex variant reflects the output of a much more recent sound change. Dating this change, however, is extremely difficult, as it would be expected to leave no orthographic evidence.

Reduction. All of the dental/alveolar realizations [r̄ r̄̄] were merged as a single variant. The uvular realization [ʁ] was excluded from consideration since, strictly speaking, it is impossible to tell whether the coronal variant which existed before /r/ backing was dental, alveolar, postalveolar or retroflex; however, since these back realizations existed in a delimited area fully embedded within the nonretroflex domain, this has no effect on the overall distribution of this change.

40. /r/ > [hr]

Variants: aspirated [hɹ], nonaspirated /r/

Reference: V.1.2 (*roof*), V.1.15 (*rubbish*), V.2.4 (*rooms*), V.10.7 (*red*), VI.6.9 (*wrist*), VI.7.15 (*reach, reached*)

Background. A small area in Somerset shows consistent aspiration of word-initial /r/. This does not result in any mergers or splits, and so is classed as a phonetic variable. We know of no evidence by which to date this change.

Reduction. Place and manner of articulation of /r/ were ignored, so that [ɹ̥ ɹ ɹ̥ ɹ̥] were all treated together as not showing aspiration.

41. Rhoticity

Variants: rhoticity, no rhoticity

Reference: II.5.1 (*corn*), II.5.1 (*barley*), V.3.1 (*fire*), V.6.2 (*barm*), V.6.7 (*burnt/burned*)

Background. This phonological variable refers to the nonrealization of etymological nonprevocalic /r/, excluding the sequence /rs/. Sporadic loss of coda /r/, especially preceding coronals, is attested from the OE period, but this is regarded as a separate sound change with different phonological consequences (contra Minkova [171]); reflexes of that earlier change are seen in variation in etymological /rs/ clusters in the SED. Evidence for the general nonprevocalic /r/ loss which we examine here becomes clear from the mid-17th [172] or around the turn of the 18th century [173], so we assume that this variable has existed in this form for around 250 years.

Reduction. All consonantal realizations of postvocalic /r/, including r-colouring of the preceding vowel, were merged as /r/ (rhoticity); all others were classed as Ø (loss of rhoticity).

42. th-fronting before /r/

Variants: fronted, nonfronted

Reference: V.10.2 (*thread*), VI.6.3 (*throat*), VII.1.3 (*three*), VIII.7.7 (*throwing*)

Background. Th-fronting refers to the sound change /θ/ > /f/; this is a phonological variable, since it results in merger with existing /f/. We have two datasets representing th-fronting, on the basis that its distribution preceding /r/ appears to be quite different to its distribution in other positions, suggesting that it represents a different change.

Reduction. Voicing was ignored, so that [f v] were merged as showing fronting, and [d, ð, ɖ, t, ʈ, ʈ^h, θ] as not showing fronting.

43. th-fronting elsewhere

Variants: fronted, nonfronted

Reference: IV.3.11 (*path*), V.7.20 (*broth*), V.8.16 (*anything*), V.10.9 (*thimble*), VI.5.6 (*tooth, teeth*), VII.6.15 (*hawing*), VII.2.11 (*both*)

Background. See th-fronting before /r/ (43).

44. th-stopping before /r/

Variants: stops, fricatives

Reference: V.10.2 (*thread*), VI.6.3 (*throat*), VII.1.3 (*three*), VIII.7.7 (*throwing*)

Background. This variable refers to the change that changes the fricative /θ/ into a stop. This is a phonological change, since it results in merger with existing coronal stop phonemes. It seems to apply regardless of other changes which affect voicing and place of articulation, so we find dental, alveolar and retroflex, voiced and voiceless variants as a result. It is likely that it postdates initial fricative voicing, so this indicates that [ð] was affected by this change just as [θ] was. The changes in place of articulation among coronal places (retroflexion before /r/, retraction to alveolar) do not apply to fricative variants, and so must postdate th-stopping. Th-fronting does not apply to the stopped variants, and so must also postdate th-stopping.

Reduction. Voicing was ignored, so that [d, ɖ, t, ʈ, ʈ^h] were merged as showing stopping and [ð, θ, v, f] as not showing stopping.

45. THOUGHT hyperrhoticity

Variants: rhoticity, no rhoticity

Reference: II.5.2 (*straw*), IV.8.8 (*straw*), V.8.9 (*draw*), VII.3.7 (*autumn*), VIII.1.4 (*daughter*), VIII.1.11 (*brought*), IX.3.8 (*caught*), IX.4.6 (*ought*), IX.4.7 (*ought*), IX.4.9 (*ought*)

Background. This variable refers to the reanalysis of some long low vowels as representing underlying /Vr/ sequences, with the result that some varieties show a rhotic realization in vowels which have no etymological *r. Here we look only at this phenomenon in the THOUGHT vowel (using the label from Wells' lexical sets [149]); this can result in merger with existing /Vr/ sequences, and so is classified as a phonological variable.

Reduction. Vowel quality was ignored, so that [aɹ, aɹ̥, ɔɹ, ɔɹ̥, ʊəɹ] were merged as showing hyperrhoticity, and [aɹ̥, æɹ̥, æ:ə, æʊ, aʊ, aɪ, αʊ, ɒ, ɒɪ, ɒʊ, ɒʊ, eɪ, eɪ, əɪ, əʊ, əʊ, eɪ, eʊ, iə, oɪ, oɪ, oɪ, oə, oʊ, ɔ, ɔ, ɔə, ɔʊ, əə, ɔʊ, u, ʊ, ʊə, ʊu, ʌʊ] as not showing it.

46. ME /ɔu/ monophthongization

Variants: monophthongized, nonmonophthongized

Reference: VIII.1.11 (*brought*), IX.4.6 (*ought*), IX.4.7 (*ought*), IX.4.9 (*ought*)

Background. This variable concerns the monophthongization of ME /ɔu/; this change did not take place in the north of England. It is classed as a phonological variable, since it results in merger with other sources of /ɔ:/ (forming the THOUGHT set).

Reduction. Among nonmonophthongized realizations, the quality of the initial element was ignored, so that [æʊ, aʊ, αʊ, ɒʊ, eʊ, oʊ, ɔʊ, ʌʊ] were merged as nonmonophthongized; various later sound changes (including later diphthongizations) were ignored, so that [aɹ, æɹ, aɹ̥, aɹ̥, ɔɹ, ɒɹ, eɹ, eɹ, oə, ɔ, ɔ, ɔə, ɔə, ɔɹ, ɔɹ̥, ʊ, ʊə] were merged as showing monophthongization.

47. wasp metathesis

Variants: wasp, waps

Reference: IV.8.7

Background. This is a phonolexical variable in the sense that it refers to a sound change which affected a single word: metathesis in the coda consonant cluster of *waps*. Both forms are attested from the OE period: Bosworth & Toller list forms *waefs*, *wæps*, *weaps* and *wæsp* [174], of which the *-fs-* forms appear to be earlier; there are just four instances of the *-sp-* forms in the Old English Web Corpus [104], and all are 10th century or later. The impression that the *-sp-* forms are the innovation is confirmed by comparison with cognates outside English, such as OHG *wefsa*. Thus the variation has existed since at least the 10th century.

Reduction. Other variants were unrelated lexical items and relatively rare, and accordingly were excluded.

48. ask metathesis

Variants: ask, aks

Reference: IX.2.4

Background. This variable deals with variation in the form of the stem of the verb *ask*. The conservative form has /sk/, as demonstrated by cognates elsewhere in West Germanic (Old Frisian *āskia*, Old Saxon *ēskon*, Old High German

eiscōn). The metathesized form has /ks/ and is abundantly attested from the OE period [175,176]; the OED suggests that in OE the metathesized form is particularly characteristic of West Saxon, consistent with the fact that the metathesized form is more common in the south and the Midlands during the ME period [177,178].

Reduction. Realizations with a postalveolar fricative ([aʃ] and similar) were merged into the nonmetathesized variant on the basis the development /sk/ > /ʃ/ is characteristic of some varieties of OE whilst */ks/ > /ʃ/ does not occur, and so a postalveolar fricative implies that metathesis never took place. Realizations with an alveolar fricative but no stop ([as] and similar) were excluded on the basis that it is impossible to tell whether these represent a reduced form of /ask/ or /aks/.

49. Birch palatalization

Variants: palatalized *birch*, nonpalatalized *birk*

Reference: IV.10.1

Background. Velar stops */k g/ underwent palatalization before the OE period, probably first becoming *[c] and *[j j̥], later undergoing lenition to [tʃ] and [j ɕ]. There appears to have been variation between palatalized and velar consonants in many words in the OE period, although OE orthography did not distinguish the palatalized and velar consonants, making it difficult to assess the situation precisely; certainly, by the ME period when the orthography begins to make these distinctions clearly, such variation is widespread, with nonpalatalized reflexes in northern and Danelaw areas, and palatalized reflexes from the south and the Midlands. We follow Ringe and Taylor [179] in assuming that this variation did not reflect dialectal variation in the application of the palatalization rule, but borrowing of cognate forms without palatalization from ON. Thus although we see apparently similar phonological variation across multiple words, this must have spread lexically; certainly changes in distribution in the following centuries has progressed differently for different stems. For these reasons, we treat palatalization in *birch* (49), *bridge* (50), *chaff* (51), and *reach* (55) as separate variables.

50. Bridge palatalization

Variants: palatalized *bridge*, nonpalatalized *brig*

Reference: IV.1.2

Background. See *birch* palatalization (49).

51. Chaff palatalization

Variants: palatalized *chaff*, nonpalatalized *caff*

Reference: II.8.5, III.5.3

Background. See *birch* palatalization (49).

52. Dove final consonant

Variants: *dove*, *doe*

Reference: IV.7.4

Background. This word has shown variation in the presence of the final consonant since the ME period; the conservative form has the consonant, the innovative form does not. The MED cites forms which probably lack final /f/ or /v/ from the first half of the 15th century onwards (*The Fire of Love* 1435, *The Book of Margery Kempe*, Book 1 1438) [180].

Reduction. Compounded variants were merged with their respective un-compounded variants: *ringdoe* with *doe*, *ringdove*, and *turtledove* with *dove*.

53. Partridge final consonant

Variants: voiced *partridge*, voiceless *partrich*

Reference: IV.7.8.(a)

Background. This word is a borrowing from French (Anglo-Norman *pardriz*, *partreiz*, Old French *perdriz*), first occurring in ME with final /tʃ/ reflecting French /ts/. The innovation is thus the voicing of the final consonant. This sound change is not restricted to this lexical item alone, but the results are not fully regular, and the variation in *partridge* does not appear to pattern with other lexical items with coda /tʃ~ɕ/ variation in the SED, so it is here treated as a phonolexical variable. On the basis of citations in the OED and MED, the voiced variant seems to have occurred from the mid-15th century (*Terms of Association* (1) 1450, *Sir Gawain and the Carl of Carlisle* 1475) [181].

54. Porridge final consonant

Variants: voiced *porridge*, voiceless *porrich*

Reference: V.7.3

Background. This variable concerns the voicing of the coda consonant of the final syllable of *porridge*. This word, originally also with different medial consonant, is a loanword from French (cf., Old French *potage*); the conservative form is thus voiced /ɕ/, and the innovation is devoicing to /tʃ/. The MED offers just one citation for the voiceless form, tentatively dated in the mid-15th century [*Herbal (misidentified as a ME version of the De Viribus Herbarum of 'Macer Floridus')* 1425(?)] [182]; the earliest attestation of the voiceless form in the OED is not until much later [183], but a voiceless final consonant is recorded for another related form, *poddish*, in the 16th century [184]. Thus we accept the dating of the innovation implied by the mid-15th century attestation.

Reduction. The two voiceless realizations, /ʃ/ and /tʃ/, were merged on the basis that it is reasonable to assume that /ʃ/ went through an earlier stage as /tʃ/.

55. Reach palatalization

Variants: palatalized *reach*, nonpalatalized *reak*

Reference: VI.7.15

Background. See *birch* palatalization (49).

56. /kw/ > /tw/

Variants: /kw/, /tw/

Reference: V.2.11 (*quilt*), VI.7.9 (*quick*)

Background. This is a phonological variable, concerning the merger of /kw/ into /tw/. This sound change is centered on the north-west of England.

Reduction. The variant [t] was merged into /tw/. Whether this just reflects a fast-speech phenomenon, or truly an additional sound change, it presupposes the application of /kw/ > /tw/.

57. BATH lengthening

Variants: short, long

Reference: I.7.7 (*shaft*), II.1.3 (*pasture*), II.8.5 (*chaff*), II.9.1 (*grass*), III.5.3 (*chaff*), III.5.4 (*grass*), IV.3.11 (*path*), IX.2.4 (*ask*)

Background. This variable concerns lengthening in ME /a/ preceding voiceless fricatives. This is generally referred to as the TRAP-BATH split, but this term is used to refer to both lengthening (whether contrastive or not) and later backing [a] > [a:] > [ɑ:]. Here we refer to it as BATH lengthening to make it clear that we are dealing only with the changes in length and not quality, since the change in length is the

earlier change. This is a phonological change since it results in a phonemic split in the /a/ vowel (unambiguously so in varieties which also undergo later backing, but, we would argue, also in varieties which do not; cf., [185]; contra [186,187]).

Reduction. All long variants [æ:, a:, a:] are merged as long; [ɑ] is also treated as long, as, since only the long vowel underwent backing, it must reflect an earlier lengthened form. All other short variants [æ, a] are merged as short.

58. CARD-CORD merger

Variants: merger, no merger

Reference: II.5.1 (*corn, barley*), V.6.2 (*barm*)

Background. This variable refers to the merger of the START and NORTH lexical sets (to use the terminology of Wells [149]), which takes place by loss of rounding of the NORTH vowel. It is a phonological change. This was assessed by comparing words with the NORTH vowel and words with the START vowel and identifying where a speaker had the same vowel for both.

59. CLOTH lengthening

Variants: lengthening, no lengthening

Reference: V.7.20 (*broth*), VII.4.7 (*on*), VII.6.6 (*frost*)

Background. The low back unrounded vowel /ɒ/ underwent lengthening in certain contexts in many varieties of English, with the result that this set merged with the THOUGHT set or, in a few cases, with the BATH set. This was assessed by comparing words with in the CLOTH set with words in the LOT, BATH, TRAP and THOUGHT sets and identifying which pair the speaker had the same vowel for. Note that this change is often referred to as the LOT-CLOTH split; we refer to it as CLOTH lengthening here to indicate that we treat speakers with same vowel in CLOTH and BATH together with speakers with the same vowel in CLOTH and THOUGHT.

60. FOOT-STRUT split

Variants: split, no split

Reference: IV.4.4 (*cut*), IV.6.14 (*ducks*), IV.6.21 (*pluck*), IV.7.4 (*doves*), IV.7.5 (*gull*), IV.9.2 (*slugs*), IV.12.4 (*stump*), V.1.15 (*rubbish*), V.2.5 (*upstairs*), V.8.4 (*some*), V.9.12 (*up*), VI.5.7 (*double*), VII.6.10 (*dull*)

Background. ME /u/ underwent loss of rounding and lowering in certain lexical items in many varieties of English, resulting in a split into the STRUT set (which undergoes the change) and the FOOT set (which does not). This is a phonological variable.

Reduction. The lowered, unrounded variants [ə, ʌ] were treated together as evidencing the split. The front rounded variant [œ:] and other minor variants, mostly reflecting lexical variation in which phoneme was found in particular words rather than variation in realization of the phoneme, were excluded. The back unrounded variant [ɜ] represents a problem. It occurs in two regions: Norfolk and Northumberland. Norfolk is mostly within the split area and Northumberland mostly within the nonsplit area, although both are adjacent to isoglosses (since Scottish English has the split, although with a rather different history). In Norfolk, speakers who use [ɜ] consistently also use [ʌ] for STRUT words and never use [ɜ] for FOOT words, suggesting [ɜ] should be treated as a variant of /ʌ/ and so as evidence for the split. In Northumberland, speakers who use [ɜ] consistently also use [ʊ] and sometimes

also use [ɜ] for FOOT words, suggesting that [ɜ] is a variant of /ʊ/ and so evidence against the split. For this reason, [ɜ] was excluded from consideration.

61. Intrusive r

Variants: intrusive r, no intrusive r

Reference: I.7.16 (*sawing-horse*), VII.6.15 (*thawing*)

Background. This variable describes the excrescence of a nonetymological /r/ to break up the hiatus between a nonhigh vowel and a following vowel. This is generally understood to be a consequence of the loss of rhoticity and so cannot be dated any later than that sound change [188]; the earliest evidence for the sound change is from Sheridan ([189] cf., [190]), so we cannot assume an age for this variable of greater than 150 years.

Reduction. All consonantal realizations of /r/ were merged as showing intrusive r, regardless of place of articulation. Instances of intrusive l were excluded from consideration.

62. ME /o:/: shortening

Variants: lax/short [ʊ, ʊ, ʌ, ʌ:, ɜ, ɜ], other

Reference: V.1.2 (*roof*), V.2.4 (*rooms*), V.2.14 (*broom*), VI.5.6 (*tooth*)

Background. See ME /o:/: EMoE fronting (32).

Reduction. Shortened monophthongs [ʊ, ʊ, ʌ, ʌ:, ɜ, ɜ] were merged as showing the change. Short [ɪ] was included in this category on the assumption that it was a variant realization of /ʊ/. Long [ʌ:] was also included in this category on the basis that it must be a lengthened realization of /ʌ/ (and thus had previously undergone shortening), in order for it to undergo the FOOT-STRUT split. All other variants [ɜ, ɜ:, ɛʊ, aʊ, ɒʊ, ou, ɔʊ, əʊ, u:, ʊu:, ʊu, ʊi, iu:, əu:, iə, iə, iʊ, iə, iu, iɜ, jɜ, jʊ] were merged as not showing the change.

63. ME /o:/: ME fronting

Variants: fronted [iə, iə, iʊ, iə, iu, iɜ, jɜ, jʊ], other

Reference: V.1.2 (*roof*), V.2.4 (*rooms*), V.2.14 (*broom*), VI.5.6 (*tooth*)

Background. See ME /o:/: EMoE fronting (32).

Reduction. Diphthongs with a close front first element [iə, iə, iʊ, iə, iu, iɜ, jɜ, jʊ] were merged as showing the change, all other variants [ɜ, ɜ:, ɛʊ, aʊ, ɒʊ, ou, ɔʊ, əʊ, u:, ʊu:, ʊu, ʊi, iu:, əu:, ɛ, ʊ, ʌ, ʌ:, ɜ] were merged as not showing the change.

64. ME /u:/: Great Vowel Shift

Variants: GVS applied to /u:/, GVS did not apply to /u:/

Reference: IV.5.2 (*mouse*), V.1.1(a) (*houses*), VI.14.14 (*trousers*), VII.1.16 (*thousand*)

Background. The SED shows a great variety of reflexes of ME /u:/. Two sound changes have been selected here as separate variables on the basis that their reflexes are clearly distinct, and occur in well-separated regions. The first is the application of the Great Vowel Shift: diphthongization of /u:/, presumably first to [əu] or something similar [191]; this is a phonological variable that dates back to the EMoE period. The second is the monophthongization of the MOUTH vowel (i.e., the reflex of ME /u:/ in those varieties in which it did undergo the GVS) to a long, low vowel (sometimes with subsequent rediphthongization with an offglide) [192]; this is a phonetic variable, since the varieties in question did not undergo the TRAP-BATH split and so had no other phonemic

long low vowel (for this latter variable, see ME /u:/: MOUTH monophthongization (33)).

Reduction. High back vowels [ɔu:, u:, u] were merged as not showing the GVS; all other reflexes [ɑ:, ea, e:, ea, eə, a:, aɪə, æ:, æ:ə, æ:ə, æa, æə, ai, ai, əu, əu:, əv, əv:, ʌv, ʊu, ʊv, eu, ʊvʊ, eu:, ɛw, eu:, ev, aʊ, au, au:, av, av, av, av, æu, æv, æ:ʊ, æɣ, æɣ:] were merged as showing the GVS.

65. *ng-coalescence*

Variants: [ŋ], [ŋg]

Reference: I.7.3 (*string*), VI.5.4 (*tongue*), IX.2.12 (*among*)

Background. This variable concerns the realization of ME /ng/ at morpheme boundaries: in some varieties the nonnasal stop element is deleted [ŋg] > [ŋ], whereas other varieties retain [ŋg]. This is a phonological change, since the result is that in varieties with the change, [ŋ] can contrast with [n] and so represents a separate (if marginal) phoneme /ŋ/.

66. */rV/ metathesis*

Variants: metathesis /Vr/, no metathesis /rV/

Reference: IV.1.2 (*bridge*), V.2.14 (*brush*), V.9.11 (*brush*), V.10.7 (*red*), VII.4.8 (*Christmas*)

Background. This variable concerns metathesis in /rV/ sequences so that *bridge* is realized [bɹ̥ːdʒ], *brush* [bɹ̥ːʃ], etc. Metatheses involving /r/ have long been a feature of the phonology of all English varieties and many alternations between /(C)VrC/ and /(C)rVC/ are dated to the OE period [193,194]. However, for the words in question, we appear to be dealing with a specific, much later and more locally delimited sound change. Of these words, *brush* has only ex-

isted in English since the ME period, and for *bridge*, *brush* and *red*, the OED and MED record no metathesized forms in ME [195]; for *Christmas* a metathesized form is recorded (*Churchwardens' Accounts of the Parish of St. Mary, Thame* 1442) [196], however, since this is not in the same region as the metathesis we see in these data, it is reasonable to assume that it is an independent sound change. The EDD records *bursh* in Somerset [197]. Thus we take this to be a recent sound change.

Reduction. Occasional realizations with a fully deleted /r/ were excluded on the basis that it is impossible to tell whether these represent metathesis followed by loss of rhoticity, or simplification of an onset /Cr/ cluster.

67. *First equative conjunction*

Variants: *so*, *as*

Reference: VIII.1.22

Background. This is a syntactic variable referring to the first conjunction used in the construction which expresses that an adjective has identical degree for two referents (e.g., *as good as* versus *so good as*). Of the two conjunctions, *as* descends from OE *ealswā* and *so* from OE *swā*; both of these are found in equative constructions already in OE (cf., examples cited in the OED and Bosworth and Toller [198,199]), so we can assume the variation is at least 1000 years old (note, however, that the *second* conjunction in this construction also differed in OE). *Swā* is far more common in this construction in OE and etymologically *ealswā* is a compounded variant of *swā*, so we can assume that *ealswā* is the innovation.

APPENDIX F: LINGUISTIC ABBREVIATIONS AND GLOSSARY

Abbreviation	
EDD	English Dialect Dictionary [200]
EME	Early Middle English
EMoE	Early Modern English
LAEME	A Linguistic Atlas of Early Middle English [106]
LALME	A Linguistic Atlas of Late Medieval English [201]
ME	Middle English
MED	Middle English Dictionary [202]
MoE	Modern English
OE	Old English
OED	Oxford English Dictionary
ON	Old Norse
PG	Proto-Germanic
Term	Gloss
1sg., 2sg., 3sg., 1pl, 2pl, 3pl	first person singular, second person singular, etc.
analogy	process of language change whereby the form of one morpheme or word influences the form of another
coda	the last part of the syllable, typically the zero or more consonants which follow the vowel
comparative	the form of an adjective which expresses higher degree for one referent relative to another
conjunction	a word that has the function of linking other words or phrases such as <i>and</i> , <i>or</i> , <i>as</i> , etc.
conservatism	the historically prior variant
conservative	(of a variant) historically prior
consonant cluster	a sequence of adjacent consonants without intervening vowel

construction	an arrangement of multiple words to express a given grammatical function
degree	the relational extent to which an adjective applies to a referent (typically distinguishing positive versus comparative versus superlative)
function words versus content words	function words are those which have little lexical meaning of their own but instead express grammatical relationships among other words in the sentence; content words are those words which do have lexical meaning
innovation	(1) the historically more recent variant; (2) the process by which a new variant is introduced to the language
innovative	(of a variant) historically more recent
isogloss	boundary between spatial domains in which different variants are dominant
levelling	(1) the spread of a single form through a paradigm so that cells which were previously morphologically distinct are no longer so; (2) the spread of a single form across communities so that there is no interspeaker variation where such variation previously existed
lexical	having to do with individual words
lexical item	a word and all its morphological forms (such as <i>speak</i> , including <i>speaks</i> , <i>speaking</i> , <i>spoke</i> , <i>spoken</i>)
metathesis	a sound change by which two phonemes in a word exchange positions
morpheme	meaningful units smaller than words, such as prefixes, suffixes and stems
morpholexical (and phonolexical, etc.)	-lexical as a suffix here indicates variation which applies only to a single word and does not reflect a wider pattern of variation in the language; for example, “morpholexical” refers to variables having to do with the formation (morpho-) of a specific word from its constituent parts where this is not part of a larger pattern involving other words
morphology, morphological	having to do with the formation of words from morphemes
nonprevocalic	not preceding a vowel (i.e., preceding a consonant or a word-boundary)
onset	the first part of the syllable, typically the zero or more consonants which precede the vowel
phoneme	a unit of sound which can distinguish words
phonetic	having to do with the realization of particular sounds where this does not affect the structure of the overall system in which those sounds are placed
phonology, phonological	having to do with the system of contrastive sounds (phonemes) used by a language to construct morphemes
preterite	past tense
prevocalic	preceding a vowel
sound change	changes in pronunciation
suppletion	a morphological property where cells in a single paradigm are supplied by unrelated stems
syntax, syntactic	having to do with the formation of sentences from words (covering word order, choice of function words, etc.)
(linguistic) variable	a linguistic context in which there is variation in form with no corresponding variation in function/meaning (“two ways of saying the same thing” [203])
variant	one possible form of a given variable

[1] R. W. Fasold and J. Connor-Linton, *An Introduction to Language and Linguistics* (Cambridge University Press, Cambridge, 2014).

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

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

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

[5] L. Bloomfield, *Language* (Holt, Rinehart and Winston, 1933).

[6] J. Milroy and L. Milroy, Linguistic change, social network and speaker innovation, *J. Linguist.* **21**, 339 (1985).

[7] V. Loreto, A. Baronchelli, A. Mukherjee, A. Puglisi, and F. Tria, Statistical physics of language dynamics, *J. Stat. Mech.* (2011) P04006.

[8] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, *Rev. Mod. Phys.* **81**, 591 (2009).

- [9] P. M. C. de Oliveira, D. Stauffer, S. Weichmann, and S. M. de Oliveira, A computer simulation of language families, *J. Linguist.* **44**, 659 (2008).
- [10] J. Burridge, Spatial Evolution of Human Dialects, *Phys. Rev. X* **7**, 031008 (2017).
- [11] J. Burridge, T. Blaxter, and B. Vaux, Evolutionary paths of language, *Europhys. Lett.* **128**, 28003 (2019).
- [12] D. M. Abrams and S. H. Strogatz, Modelling the dynamics of language death, *Nature (London)* **424**, 900 (2003).
- [13] G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane, Utterance selection model of language change, *Phys. Rev. E* **73**, 046118 (2006).
- [14] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, Sharp transition towards shared vocabularies in multi-agent systems, *J. Stat. Mech.* (2006) P06014.
- [15] A. Baronchelli, L. Dall'Asta, A. Barrat, and V. Loreto, Topology-induced coarsening in language games, *Phys. Rev. E* **73**, 015102 (2006).
- [16] P. A. P. Moran, Random processes in genetics, *Proc. Cambridge Philos. Soc.* **54**, 60 (1958).
- [17] R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford, 1930).
- [18] M. Kimura and G. H. Weiss, The stepping stone model of population structure and decrease of genetical correlation with distance, *Genetics* **49**, 561 (1964).
- [19] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).
- [20] P. L. Krapivsky, S. Redner, and Eli Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, 2010).
- [21] J. T. Cox and R. Durrett, The stepping stone model: New formulas expose old myths, *Ann. Appl. Probab.* **12**, 1348 (2002).
- [22] A. Etheridge, *Some Mathematical Models from Population Genetics* (Springer, London, 2009).
- [23] R. A. Blythe and A. J. McKane, Stochastic models of evolution in genetics, ecology and linguistics, *J. Stat. Mech.* (2007) P07018.
- [24] H. Kauhanen, Neutral change, *J. Linguist.* **53**, 327 (2017).
- [25] J. Burridge and S. Kenney, Birdsong dialect patterns explained using magnetic domains, *Phys. Rev. E* **93**, 062402 (2016).
- [26] T. Tesileanu and H. Meyer-Ortmanns, Competition of languages and their hamming distance, *Int. J. Modern Physics C* **17**, 259 (2006).
- [27] J. Burridge, Unifying models of dialect spread and extinction using surface tension dynamics, *R. Soc. opensci.* **5**, 171446 (2018).
- [28] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996).
- [29] A. J. Bray, Theory of phase-ordering kinetics, *Adv. Phys.* **43**, 357 (1994).
- [30] R. J. Glauber, Time-dependent statistics of the ising model, *J. Math. Phys.* **4**, 294 (1963).
- [31] I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen, Critical Coarsening Without Surface Tension: The Universality Class of the Voter Model, *Phys. Rev. Lett.* **87**, 045701 (2001).
- [32] J.-M. Drouffe and C. Godreche, Phase ordering and persistence in a class of stochastic processes interpolating between the ising and voter models, *J. Phys. A: Math. Gen.* **32**, 249 (1999).
- [33] P. Chen and S. Redner, Majority rule dynamics in finite dimensions, *Phys. Rev. E* **71**, 036101 (2005).
- [34] S. Asch, Studies of independence and conformity: i. a minority of one against a unanimous majority, *Psychol. Monogr.: Gen. Appl.* **70**, 1 (1956).
- [35] 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).
- [36] T. M. Liggett, *Interacting Particle Systems* (Springer, New York, 1985).
- [37] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [38] J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA* **81**, 3088 (1984).
- [39] M. A. Nowak, N. L. Komarova, and P. Niyogi, Evolution of universal grammar, *Science* **291**, 114 (2001).
- [40] W. G. Mitchener, A mathematical model of prediction-driven instability: How social structure can drive language change, *J. Logic Lang. Inf.* **20**, 385 (2011).
- [41] W. G. Mitchener, A stochastic model of language change through social structure and prediction driven instability, *SIAM J. Appl. Math.* **77**, 2272 (2017).
- [42] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, 2003).
- [43] H. Orton and E. Dieth, *Survey of English Dialects* (E. J. Arnold, Leeds, 1962).
- [44] A. Leemann, M.-J. Kolly, R. Purves, D. Britain, and E. Glaser, Crowdsourcing language change with smartphone applications, *PLoS ONE* **11**, 1 (2016).
- [45] B. Vaux and M. L. Jøhndal, Cambridge survey of world englishes, 2017.
- [46] D. A. Ringe and J. F. Eska, *Historical Linguistics: Toward a Twenty-First Century Reintegration* (Cambridge Univ. Press, Cambridge, 2013).
- [47] W. Labov, *Principles of Linguistic Change, Vol. 1: Internal Factors of Language in Society*, (Blackwell, Oxford, UK; Cambridge [Mass.], 1994).
- [48] G. Sankoff, Language Change Across the Lifespan, *Ann. Rev. Linguist.* **4**, 297 (2018).
- [49] H. Raumolin-Brunberg, Lifespan changes in the language of three early modern gentlemen, *The Language of Daily Life in England (1450–1800)*, edited by Minna Nevala, Arja Nurmi, and Minna Palander-Collin (John Benjamins, Amsterdam, 2009).
- [50] C. L. H. Kam and E. L. Newport, Regularizing unpredictable variation: The roles of adult and child learners in language formation and change, *Lang. Learn. Dev.* **1**, 151 (2005).
- [51] J. Mathews and R. L. Walker, *Mathematical Methods of Physics* (Addison Wesley, 1971).
- [52] E. A. Wrigley and R. S. Schofield, *The Population History of England 1541-1871* (Cambridge University Press, Cambridge, 1989).
- [53] D. Crystal, *The Cambridge Encyclopedia of the English Language* (Cambridge University Press, Cambridge, 2003).
- [54] G. Grimmett and D. Stirzaker, *Probability and Random Processes* (Oxford University Press, Oxford, 2001).

- [55] M. Kimura and Maruyama, Pattern of neutral polymorphism in a geographically structured population, *Genet. Res.* **18**, 121 (1971).
- [56] J. Séguy, La relation entre la distance spatiale et la distance lexicale, *Revue de linguistique romane* **35**, 335 (1971).
- [57] J. Séguy, La dialectométrie dans l'atlas linguistique de la Gascogne, *Revue de linguistique romane* **37**, 1 (1973).
- [58] J. Nerbonne and P. Kleiweg, Lexical distance in lamsas, *Comp. Humanit.* **37**, 339 (2003).
- [59] J. Grieve, D. Speelman, and D. Geeraerts, A statistical method for the identification and aggregation of regional linguistic variation, *Lang. Var. Change* **23**, 193 (2011).
- [60] P. Moran, Notes on continuous stochastic phenomena, *Biometrika* **37**, 17 (1950).
- [61] M. W. Beresford and J. K. S. St. Joseph, *Medieval England: an Aerial Survey* (Cambridge University Press, Cambridge, 1979).
- [62] L. Wasserman, *All of Statistics* (Springer, Berlin, 2003).
- [63] D. Koller and N. Friedmann, *Probabilistic Graphical Models* (The MIT Press, Cambridge, Massachusetts, 2009).
- [64] J. Besag, Statistical analysis of non-lattice data, *J. Roy. Stat. Soc. Ser. D* **24**, 179 (1975).
- [65] R. M. Neal, Annealed importance sampling, *Stat. Comput.* **11**, 125 (2001).
- [66] L. Campbell, *Historical Linguistics* (Edinburgh University Press, Edinburgh, 2013).
- [67] C. Darwin, *The Descent of Man* (John Murray, London, 1897).
- [68] R. A. Blythe, Neutral evolution: A null model for language dynamics, *Adv. Complex Syst.* **15**, 1150015 (2012).
- [69] H. Georgii, *Stochastics: Introduction To Probability And Statistics* (De Gruyter, 2012).
- [70] R. Salakhutdinov, Learning and evaluating boltzmann machines, Technical Report UTML TR 2008-002, Dept. of Computer Science, University of Toronto, 2008.
- [71] H. Orton and E. Dieth, *Survey of English Dialects* (E. J. Arnold, Leeds, 1962).
- [72] *The Linguistic Atlas of England*, edited by H. Orton, S. Sanderson, and J. D. A. Widdowson (Croom Helm [u.a.], London, 1978), OCLC: 5029407.
- [73] C. Upton and J. D. A. Widdowson, *An Atlas of English Dialects* (Oxford University Press, New York, 1996).
- [74] International Phonetic Association, editor, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (Cambridge University Press, Cambridge, UK; New York, NY, 1999).
- [75] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth* (Clarendon Press, Oxford, 1898), *nædre*.
- [76] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, in *Middle English Compendium*, edited by F. McSparran, P. Schaffner, J. Latta, M. Logarbo, and R. E. Lewis (University of Michigan Press); G. Cumberledge, *Middle English Compendium* (Oxford University Press, Ann Arbor, Michigan; London, 1952), *hagworm* n.
- [77] B. Mitchell, *Old English Syntax Volume 1: Concord, the Parts of Speech, and the Sentence* (Clarendon Press, Oxford, 1985), §378–380.
- [78] B. Mitchell, *Old English Syntax Volume 1: Concord, the Parts of Speech, and the Sentence* (Clarendon Press, Oxford, 1985), §434.
- [79] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth* (Clarendon Press, Oxford, 1898).
- [80] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth* (Clarendon Press, Oxford, 1898).
- [81] D. Roehrs and C. Sapp, Complex quantifiers with genitive and concord in Old English and beyond, *J. Comp. Germanic Ling.* **21**, 381 (2018).
- [82] M. Markus, EDD Online 3.0, <http://eddonline-proj.uibk.ac.at/edd/index.jsp>, 2019, *ORT*, sb and v.
- [83] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *FÝST*.
- [84] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *nēve* n.(2).
- [85] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *FROGGA*.
- [86] T. Hunt, *Plant Names of Medieval England*, D. S. Brewer (Cambridge; Wolfeboro, N. H., USA, 1989).
- [87] M. Markus, EDD Online 3.0, <http://eddonline-proj.uibk.ac.at/edd/index.jsp>, 2019, *Jacky*, sb.adj. *Jacob*, sb.
- [88] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *irchōun* n.
- [89] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *heġġe* n.
- [90] J. Bosworth and T. N. Toller, *An Anglo-Saxon dictionary: Based on the manuscript collections of the late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *ādexe*; *efete*.
- [91] T. Wright, *Anglo-Saxon and Old English Vocabularies*, (Trübner and Co., London, 1884), p. 122.
- [92] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *úle*.
- [93] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *hōulot* n.
- [94] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *tēmen* v.(2).

- [95] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *pöuren* v.(2).
- [96] F. Bacon, *Sylva Sylvarum; or, A Naturall Historie*, J. H. for W. Lee, London, 1st edition, 1626.
- [97] M. Markus, EDD Online 3.0, <http://eddonline-proj.uibk.ac.at/edd/index.jsp>, 2019, *OVER*, prep, adv, adj and v.
- [98] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *vinegre* n.
- [99] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *äl ēgre, aleger* n.
- [100] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *gist; beorma*.
- [101] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *yžst* n.; *berm(e)* n.
- [102] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *brennan*.
- [103] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *barman*.
- [104] A. D. Healey, J. P. Wilkin, and X. Xiang, Dictionary of Old English Web Corpus, <https://tapor.library.utoronto.ca/doecorpus/>, 2009.
- [105] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *brennan*.
- [106] M. Laing, A Linguistic Atlas of Early Middle English, 1150–1325, <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>, 2013.
- [107] Anon, *Remarks on the English Woollen Manufactory for Exportation, and Necessity of Preventing Irish Wool Being Run*, printed for (Geo Grierson, London, 1730), p. 8.
- [108] E. Smith, *The Cure of Deism: Or the Mediatorial Scheme by Jesus Christ The Only True Religion. In Answer to the Objections Started, and to the Very Imperfect Account of The Religion of Nature, and of Christianity; given by the Two Oracles of Deism, the Author of Christianity as Old as the Creation; and the Author of the Characteristics*, volume I, Printed for the author; And sold by W. Innys and R. Manby, London, second edition, 1737, p. 349.
- [109] J. Ward, *Four Essays on the English Language*, printed for John Ward, London, 1758, p. 207.
- [110] J. J. Rousseau, *Emilius; or, an Essay on Education*, volume I, printed for J. Nourse and P. Vaillant, London, 1763, trans. Mr. Nugent. p. 291.
- [111] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *gēten* v.(1).
- [112] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *grouen* v.
- [113] R. Boyle, *The Origine of Formes and Qualities*, printed by H. Hall printer to the University, for Ric. Davis, Oxford, 1st edition, 1666, vii. iii. p. 131.
- [114] R. Plot, *The Natural History of Oxford-Shire. By R.P.*, printed at the Theater, and are to be had there, Oxford &c., 1677, v. p. 107.
- [115] G. Sinclair, *Natural Philosophy Improven by New Experiments, Touching the Mercurial Weather-Glass, the Hygroscope [&c.] By G.S.* [publisher not identified], Edinburgh, 1683, p. 4.
- [116] Á. B. Magnússon, *Íslensk orðsifjabók*, (Orðabók Háskólans: Mál og menning, Reykjavík, December 1989), *verri*.
- [117] J. Mirk, *The Festyuall*, S. L. : At the sygne of the Sonne, by me Wynkyn de Worde, London, new edition, May 1508, f. xlii^v.
- [118] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *wer* adj. (comparative).
- [119] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *hiren* pron., *hires* pron.
- [120] C. L. Allen, *Genitives in Early English: Typology and Evidence*, (Oxford University Press, Oxford; New York, 2008), p. 146.
- [121] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *hé*.
- [122] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *hisen*.
- [123] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan

- Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *oures, ouren*.
- [124] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *theires*.
- [125] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *theiren*.
- [126] K. Brunner, *Altenglische Grammatik (nach der angelsächsischen Grammatik von Eduard Sievers neubearbeitet)*, (Max Niemeyer, Tübingen, 1965), pp. 271–273.
- [127] T. Nevalainen and H. Raumolin-Brunberg, The third-person singular *-es* and *-eth* revisited: The morphophonemic hypothesis, In Christine Dalton-Puffer and Nicholas Ritt, editors, *Words: Structure, Meaning, Function: A Festschrift for Dieter Kastovsky*, pp. 235–248. (Mouton de Gruyter, Berlin, 2000), p. 242.
- [128] G. D. Miller, The origin and diffusion of English 3sg-s, *Studia Anglica Posnaniensia* 38, 353 (2002).
- [129] L. Wright, Some more on the history of present-tense *-s*, *do* and *zero*: West Oxfordshire, 1837, *Journal of Historical Sociolinguistics*, 1(1), January 2015.
- [130] B. Mitchell, *Old English Syntax Volume 1: Concord, the Parts of Speech, and the Sentence*, (Clarendon Press, Oxford, 1985), §651-664.
- [131] M. Laing, A Linguistic Atlas of Early Middle English, 1150–1325, <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>, 2013, map 12404002.
- [132] M. Benskin, V. Karaiskos, K. Williamson, and M. Laing, *An Electronic Version of A Linguistic Atlas of Late Medieval English*, (University of Edinburgh, 2013), item 178-40.
- [133] M. Laing, A Linguistic Atlas of Early Middle English, 1150–1325, <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>, 2013, *bēn* v.
- [134] M. Benskin, V. Karaiskos, K. Williamson, and M. Laing, *An Electronic Version of A Linguistic Atlas of Late Medieval English*, (University of Edinburgh, 2013), item 178-40.
- [135] M. Laing, A Linguistic Atlas of Early Middle English, 1150–1325, <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>, 2013, text #296.
- [136] N. K. de Haas, *Morphosyntactic Variation in Northern English: The Northern Subject Rule, Its Origins and Early History*, Ph.D., Radboud Universiteit, Nijmegen, 2011, pp. 81–116.
- [137] N. K. de Haas, *Morphosyntactic Variation in Northern English: The Northern Subject Rule, Its Origins and Early History*, Ph.D., Radboud Universiteit, Nijmegen, 2011, p. 82.
- [138] T. Nevalainen, H. Raumolin-Brunberg, J. Keränen, M. Nevala, A. Nurmi, M. Palander-Collin, Ann Taylor, Susan Pintzuk, and Anthony Warner, *Parsed Corpus of Early English Correspondence (PCEEC)*, (Oxford Text Archive, 2006).
- [139] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *hit* pron.
- [140] Horace, *A Medicinable Morall, That Is, the Two Bookes of Horace His Satyres, Englyshed Accordyng to the Prescription of Saint Hierome*, Imprinted by Thomas Marshe, Fletestrete, 1566, trans. T. Drant. v.
- [141] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden, Quantitative Analysis of Culture Using Millions of Digitized Books, *Science* 331, 176 (2011).
- [142] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 128–132.
- [143] W. Grant and D. D. Murrison, editors, *The Scottish National Dictionary*, (Scottish National Dictionary Association, December 1931-1936).
- [144] M. Benskin, V. Karaiskos, K. Williamson, and M. Laing, *An Electronic Version of A Linguistic Atlas of Late Medieval English*, (University of Edinburgh, 2013), item 313.
- [145] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), pp. 53–60.
- [146] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), p. 61.
- [147] A. Ellis, *On Early English Pronunciation, with Especial Reference to Shakspeare and Chaucer. Part V*, (Trübner and Co., London, 1889), p. 226.
- [148] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 104–108.
- [149] J. C. Wells, *Accents of English*, Cambridge University Press, Cambridge [Cambridgeshire]; New York, 1982.
- [150] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), pp. 89, 98, 99.
- [151] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), pp. 31, 89, 94–95.
- [152] P. M. Anderson, *A Structural Atlas of the English Dialects*, Croom Helm, London; New York, 1987, pp. 89, 93.
- [153] R. Lass, Old English fricative voicing unvisited, *Studia Anglica Posnaniensia*, XXV-XXVII:3–45, January 1991, pp. 24–26.
- [154] J. Fisiak, The voicing of initial fricatives in Middle English, *Focus on England and Wales*, 4:3–21, 1985, p. 5.
- [155] R. Lass, Old English fricative voicing unvisited, *Studia Anglica Posnaniensia*, XXV-XXVII:3–45, January 1991, p. 23.
- [156] J. Fisiak, The voicing of initial fricatives in Middle English, in *Focus on: England and Wales*, edited by W. Viereck (John Benjamins, Amsterdam, 1985).
- [157] K. Brunner, *Altenglische Grammatik (nach der angelsächsischen Grammatik von Eduard Sievers neubearbeitet)* (Max Niemeyer, Tübingen, 1965).
- [158] R. Berndt, *Einführung in das Studium des Mittlenglischen*, (Niemeyer, Halle/S., 1960).
- [159] H. E. Pinsker, *Historische englische Grammatik* (Hueber, Munich, fourth edition, 1974).
- [160] W. H. Bennet, The southern English development of Germanic initial [f s þ], *Language* 31, 367 (1955).
- [161] R. Lass, Old English fricative voicing unvisited, *Studia Anglica Posnaniensia* XXV-XXVII, 3 January (1991).

- [162] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 146–147.
- [163] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 116–118.
- [164] H. Jones, *Accidence of the English Tongue*, Number 22 in Reprint. (Scholar Press, Meston, 1976), 1724.
- [165] K. Wales, *Northern English: A Social and Cultural History*, (Cambridge University Press, Cambridge, 2006), p. 101.
- [166] H. L. C. Tristram, Linguistic contacts across the English Channel: The case of Breton retroflex ⟨r⟩, In *Linguistic Change under Contact Conditions*, pp. 291–313. (M.de Gruyter, Berlin; New York, 1995).
- [167] B. Erickson, On the development of English *r*, In Donka Minkova and Robert P. Stockwell, editors, *Studies in the History of the English Language: A Millennial Perspective*, number 39 in Topics in English Linguistics, pp. 183–206 (Mouton de Gruyter, Berlin; New York, 2002).
- [168] H. L. C. Tristram, Linguistic contacts across the English Channel: The case of Breton retroflex ⟨r⟩, In *Linguistic Change under Contact Conditions*, pp. 291–313 (M. de Gruyter, Berlin; New York, 1995), p. 303.
- [169] O. Scheiders, *Die englischen r-Isophonen - Skandinavische Substrateinwirkung?* (MA, University of Freiburg, Freiburg, 1991).
- [170] M. Trautmann, Besprechung einiger Schulbücher nebst bemerkungen über die r-laute, *Anglia* 3, 204 (1880).
- [171] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 123–124.
- [172] A. M. S. McMahon, *Lexical Phonology and the History of English*, Number 91 in Cambridge Studies in Linguistics. (Cambridge University Press, Cambridge; New York, 2000), pp. 237–241.
- [173] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 125–127.
- [174] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, Clarendon Press, Oxford, 1898, *wæps*.
- [175] J. Bosworth and T. N. Toller, *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*, (Clarendon Press, Oxford, 1898), *acsian, axian*.
- [176] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *asken*.
- [177] M. Laing, A Linguistic Atlas of Early Middle English, 1150–1325, <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>, 2013, map 27620802, map 27620804.
- [178] M. Benskin, V. Karaiskos, K. Williamson, and M. Laing, *An Electronic Version of A Linguistic Atlas of Late Medieval English*, (University of Edinburgh, 2013), item 80.
- [179] D. A. Ringe and A. A. Taylor, *A Linguistic History of English*, Oxford Linguistics. (Oxford University Press, Oxford; New York, 2006), p. 204.
- [180] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *dōuve* n.
- [181] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *partrich(e)* n.
- [182] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *potāġe* n.
- [183] J. Learmont, *Poems Pastoral, Satirical, Tragic and Comic*, (Alexander Chapman and Company, Edinburgh, 1791), p. 65.
- [184] W. Tyndale, *The Obedyence of a Chrysten Man: The Obedyence of All Degrees, Proved by Goddes Worde and Fyrst of Chylderne Unto Theyr Elders*, (Hans Lust, Marlborow, 1528), f. cxxx.
- [185] T. Blaxter and R. Coates, The trap–bath split in Bristol English, *English Language and Linguistics* 23, 1 (2019).
- [186] A. Hughes, P. Trudgill, and D. Watt, *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in the British Isles*, Hodder Arnold, London, 5th edition edition, 2012, OCLC: 253820559.
- [187] M. W. H. Kester, *The Bristol Dialect: A Comparison between Age-Groups*, Ph.D. thesis, University of Bristol, Bristol, 1979.
- [188] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 127–128.
- [189] T. Sheridan, *A Course of Lectures on Elocution*, (Strahan, London, 1762).
- [190] M. Sósokuthy, Analogy in the emergence of intrusive- *r* in English, *English Language and Linguistics*, 17(1):55–84, March 2013, p. 58.
- [191] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), p. 41.
- [192] P. M. Anderson, *A Structural Atlas of the English Dialects*, (Croom Helm, London; New York, 1987), pp. 41, 54–55.
- [193] D. A. Ringe and A. A. Taylor, *A Linguistic History of English*, Oxford Linguistics. (Oxford University Press, Oxford; New York, 2006), pp. 340–341.
- [194] D. Minkova, *A Historical Phonology of English*, Edinburgh Textbooks on the English Language - Advanced. (Edinburgh Univ. Press, Edinburgh, 2014), pp. 119–120.
- [195] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *rēd* adj.; *brīġġe* n.; *brushe* n.
- [196] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, Middle English Dictionary, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis,

- editors, *Middle English Compendium*, University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952, *Criste-mas*(se n).
- [197] F. T. Elworthy, *The West Somerset Word-Book a Glossary of Dialectal and Archaic Words and Phrases Used in the West of Somerset and East Devon* (English Dialect Society, London, 1888).
- [198] J. Bosworth and T. N. Toller, *An Anglo-Saxon dictionary: Based on the manuscript collections of the late Joseph Bosworth: Supplement* (Clarendon Press, Oxford, 1921), *swá*.
- [199] J. Bosworth and T. N. Toller, *An Anglo-Saxon dictionary: Based on the manuscript collections of the late Joseph Bosworth: Supplement* (Clarendon Press, Oxford, 1921), *eall-swá*.
- [200] M. Markus, EDD Online 3.0, <http://eddonline-proj.uibk.ac.at/edd/index.jsp>, 2019.
- [201] M. Benskin, V. Karaiskos, K. Williamson, and M. Laing, *An Electronic Version of A Linguistic Atlas of Late Medieval English* (University of Edinburgh, 2013).
- [202] R. E. Lewis, M. J. Williams, H. Kurath, S. M. Kuhn, and J. Reidy, *Middle English Dictionary*, In Frances McSparran, Paul Schaffner, John Latta, Mona Logarbo, and Robert E. Lewis, editors, *Middle English Compendium*. University of Michigan Press; G. Cumberledge, Oxford University Press, Ann Arbor, Michigan; London, January 1952.
- [203] W. Labov, *Sociolinguistic Patterns*, (University of Pennsylvania Press, Philadelphia, 1972), p. 271.