

A Model of Norm Emergence and Innovation in Language Change

Samarth Swarup
Network Dynamics and
Simulation Science Lab,
Virginia Bioinformatics
Institute,
Virginia Tech,
Blacksburg, VA, USA
swarup@vbi.vt.edu

Andrea Apolloni
Institut des Systèmes
Complexes Rhône-Alpes,
and Laboratoire de Physique,
École Normale Supérieure de
Lyon
69007 Lyon, France
andrea.apolloni@ens-
lyon.fr

Zsuzsanna Fagyal
School of Literatures,
Cultures, and Linguistics,
Department of French,
University of Illinois at
Urbana-Champaign
Urbana, IL, USA
zsfagyal@illinois.edu

ABSTRACT

We analyze and extend a recently proposed model of linguistic diffusion in social networks, to analytically derive time to convergence, and to account for the innovation phase of lexical dynamics in networks. Our new model, the degree-biased voter model with innovation, shows that the probability of existence of a norm is inversely related to innovation probability. When the innovation rate in the population is low, variants that become norms are due to a peripheral member with high probability. As the innovation rate increases, the fraction of time that the norm is a peripheral-introduced variant and the total time for which a norm exists at all in the population decrease. These results align with historical observations of rapid increase and generalization of slang words, technical terms, and new common expressions at times of cultural change in some languages.

Categories and Subject Descriptors

J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms

Algorithms, Experimentation, Theory

Keywords

Social Simulation, Lexical Innovation, Norms, Degree-biased Voter Model

1. INTRODUCTION

Multiagent modeling and analysis is being increasingly applied to the study of language change [1; 3, e.g.]. In this view, a language is seen as an emergent phenomenon from the interactions between a population of communicating agents, and change in language is driven by linguistic factors, such as frequency of use, and social factors like social

Cite as: A Model of Norm Emergence and Innovation in Language Change, Samarth Swarup, Andrea Apolloni, and Zsuzsanna Fagyal, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 693-700.

Copyright © 2011, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

network structure and popularity. Computational modeling is especially relevant with respect to language change; it provides tools to explore the large-scale consequences of small incremental changes that are typically studied empirically at the individual level or the level of small communities.

One of the foundational questions in this respect is, how do linguistic norms emerge, and how do they change? Recently, Fagyal et al. [9] proposed a model, known as the degree-biased voter model (DBVM), to study the role of network structure and popularity in the spread of linguistic variants. They showed that the DBVM brings together in one model two separate factors in the emergence of linguistic norms: the role of network positions, in particular the contribution of central and peripheral agents referred to as *hubs* and *loners*, and the role of popularity in determining which linguistic variants are preferentially copied and propagated. These factors had been separately attested to in the empirical sociolinguistic literature [5; 16; 18; 19, e.g.], but never combined into a model of norm emergence before.

However, their model left some important questions unaddressed:

- From an analytical perspective, how long does it take for a norm to emerge, i.e., what is the *time to convergence*?
- From a sociolinguistic perspective, their model does not address the *innovation phase* of the dynamics. Who creates the new variants that go on to become norms?

The first question is relevant in that the time to convergence is directly related to the time it takes to switch between norms, which would allow to investigate cycles of fashion and fad quantitatively. The second question is important for understanding diffusion dynamics at times of increased cultural contact when the innovation rate, for instance in the lexicon via borrowing or other means, is particularly high.

In the present work, we analyze the DBVM to derive time to convergence in terms of the size of the network. We also introduce an extension to the model to include innovation, and we numerically address the question of which network positions have an advantage in terms of generating new norms. Languages tend to be stable for long periods, and then change in bursts (typically triggered by large-scale social change). Our analysis and extension here, therefore,

combine to present a more complete model of linguistic dynamics. The extended model shows that in situations where the innovation rate is high, there tend to be multiple variants in competition in the network, and the period of time for which norms exist is lower. Further, many peripheral variants become the norm in the network. This behavior is congruent with qualitative observations of certain types of lexical change in French and possibly in other languages.

The rest of this paper is organized as follows: first we provide some linguistic context and a description of the DBVM, then we analyze the model to derive expressions for time to convergence. This is followed by a discussion of the DBVM dynamics, which we extend to include a parameter for innovation. We present some simulations to analyze this extended model, and show that at low innovation rates loners, i.e., peripheral agents who might influence others but do not listen to anyone else in the network, are more likely to produce the variants that later become norms in the network. As the innovation rate increases, however, both the fraction of time that the norm is a peripheral-introduced variant and the total time for which a norm exists at all in the population decrease. We discuss the relevance of this model to changes in French in the 19th Century.

2. THE DEGREE-BIASED VOTER MODEL

The importance of the social network in language change has been recognized for a long time. Bloomfield first suggested a thought experiment, where “every time any speaker uttered a sentence, an arrow were drawn into the chart pointing from his dot to the dot representing each one of his hearers. At the end of a given period of time, say seventy years, this chart would show us the density of communication within the community” [4, p. 46]. He hypothesized that these “lines of communication” and “the relative prestige of social groups” were the two main conditioning factors of “the spread of linguistic features” [4, p. 345].

Since then several researchers have studied the role of social networks in language change, by mapping out specific networks and recording the spread of linguistic variants and emergence of norms over these networks [7, 8, 15, 18, 19, 27]. Theorizing has focused on the roles of central “hubs” or “leaders” and peripheral “loners” or “lames” in the diffusion process.

These and other studies have resulted in two competing models of language change: Labov’s and Eckert’s work has supported the so-called *two-step flow of influence* model [14]. This model says that the centrally-connected leaders are responsible for introducing new variants into the local network, and that they themselves are primarily influenced by other leaders. On the other hand, work by the Milroys supports the *weak-tie model of influence* [11, 12]. In this model, it is the loosely-connected peripherals who introduce new variants into the local network, which they are able to do because they are relatively free from the regulatory influence of the local leaders, and more in touch with outsiders.

These models are at odds with each other because they posit different roles for the central and peripheral members of the network: hubs are considered agents of innovation in one, and conservative regulators in the other; peripherals are considered barely involved in the linguistic life of the network in one, and sources of novel variants in the other. The question then arises, how can these seemingly mutually contradictory explanations be reconciled?

Fagyal et al. proposed the degree-biased voter model to answer this question [9]. In this model, each node in the network (corresponding to an agent) is initialized with one variant of a linguistic variable. A variant can be phonetic, such as a flapped or fully released /t/ in the word “mittens” (number of variants, $v = 2$), or it can be a stylistic or contextual variant of a lexical item such as the French *voiture* as ‘véhicule’, ‘char’, ‘tacot’, or ‘bagnole’ ($v = 4$), etc. Further, edges in the network are directed, and an edge from node A to node B is interpreted to mean that A can copy B .

Once the simulation starts, at each time step, an agent copies a neighbor’s variant with probability proportional to the neighbor’s in-degree (the number of edges pointing to the neighbor). Thus, the probability that neighbor i will be chosen to copy from is,

$$P(i) = \frac{k_i^{in}}{\sum_j k_j^{in}}, \quad \forall i, j \in \mathcal{N} \quad (1)$$

where k_i^{in} is the in-degree of neighbor i , and \mathcal{N} is a set consisting of all the neighbors of the current node. Note that the sum in the denominator is taken over all the neighbors of the node.

They showed that on a scale-free network with a small number of loners, this model results in the rapid emergence of norms, where nearly all the agents are in the same state (except the loners initialized in a different state). Loners remain fixed in their initial states because they have no links pointing to another agent (meaning they do not copy anyone else), but can still influence the dynamics within the network because they have (a very small number of) links pointing to them (meaning others can copy them). The presence of these loners makes the system a *driven*, or out-of-equilibrium, system. Thus the norms, while stable for long periods, will eventually be replaced by other norms, as some of the agents stochastically copy one of the loners in a different state, and this new variant gets propagated through the network. Interestingly, they showed that norms do not appear if degree-biasing is not present, which implies that norms emerge only when the system is *close* to equilibrium.

Their model points to a resolution of the debate over the two competing models of language change formulated by linguists by suggesting that both interpretations can be seen as valid at different instants of observation of the stochastic process of linguistic diffusion. Hubs essentially fulfill the roles of enforcing norms, but they also rapidly spread new variants when they themselves change their state. Loners tend to hold on to their variants, which then sometimes stochastically work their way up to the hubs because of short path lengths in a scale-free network, and thereby trigger changes in norms.

In this paper, we make this analysis more quantitative by analytically deriving the time-scale of norm emergence, as we now do.

3. ANALYZING THE DBVM

For analysis, we simplify the model slightly, by considering a system of N nodes connected through *undirected* links. We indicate with k the degree of each node and with n_k the fraction of nodes with degree k . We suppose that the degree distribution is a power law with exponent ν .

We also assume that the network is perfectly uncorrelated (a Molloy-Reed network [20]), which means that the proba-

bility of an edge between any two nodes is given by,

$$P(\text{edge } xy) = \frac{k_x k_y}{N^2}.$$

Thus, the probability that node x copies node y in the DBVM is given by,

$$\begin{aligned} P(x \text{ copies } y) &= P(\text{edge } xy) \frac{k_y^\beta}{\sum_j P(\text{edge } xj) k_j^\beta}, \\ &= \frac{\frac{k_x k_y}{N^2} k_y^\beta}{\frac{k_x}{N^2} \sum_j k_j k_j^\beta}, \\ &= \frac{k_y^{\beta+1}}{\sum_j k_j^{\beta+1}}, \end{aligned}$$

where the summation is over all the nodes in the network. The coefficient β is the weight of the node. When $\beta = 0$ we obtain the standard voter model [23, 24], and when $\beta = 1$, we obtain the canonical DBVM. Now, $\sum_j k_j^{\beta+1} = N \sum_k k^{\beta+1} n_k$, where n_k is the fraction of nodes of degree k . We define $\mu_{\beta+1} = \sum_k k^{\beta+1} n_k$ as momentum of order $\beta + 1$. Therefore,

$$P(x \text{ copies } y) = \frac{k_y^{\beta+1}}{N \mu_{\beta+1}}. \quad (2)$$

We further assume that a node can have one of only two variants or states (i.e., $v = 2$), which we denote with $+1$ (state up) or -1 (state down). We indicate with ρ_k (correspondingly, $1 - \rho_k$) the fraction of nodes with degree k in state up (state down). At each iteration a node is chosen and one of its neighbours is picked up: if the states of the two nodes are different the first node copies the state of the second one with a probability based on the degree of the second one. The probability for a node with degree k and state down to switch state can be shown to be given by:

$$R_k(\rho_k) = n_k(1 - \rho_k) \sum_j \frac{j^{\beta+1} n_j \rho_j}{\mu_{\beta+1}} = n_k(1 - \rho_k) \omega_{\beta+1} \quad (3)$$

where $\omega_{\beta+1}$ is called the weighted magnetization. Correspondingly, the probability for a node with degree k and state up to switch is given by:

$$L_k(\rho_k) = n_k \rho_k \sum_j \frac{j^{\beta+1} n_j (1 - \rho_j)}{\mu_{\beta+1}} = n_k \rho_k (1 - \omega_{\beta+1}). \quad (4)$$

From now on we concentrate on the DBVM, which mean we assume $\beta = 1$ in what follows. The state of the system is defined at every time by the vector $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_k)$ representing the fraction of nodes with degree k and state $+1$. We indicate with $P(\boldsymbol{\rho}, t)$ the probability that the system at time t is in the configuration $\boldsymbol{\rho}$. At each time step, the fraction ρ_k can change by a quantity $\delta_k = \frac{1}{N n_k}$ representing the fact that one of the nodes has switched state. Indicating with $\delta t = 1/N$, the time evolution of the system is ruled by:

$$\begin{aligned} P(\boldsymbol{\rho}, t + \delta t) &= P(\boldsymbol{\rho}, t) + \sum_k L_k(\rho_k + \delta_k) P(\rho_k + \delta_k, t) \\ &+ \sum_k R_k(\rho_k - \delta_k) P(\rho_k - \delta_k, t) \\ &- \sum_k (R_k(\rho_k) + L_k(\rho_k)) P(\rho_k, t) \end{aligned} \quad (5)$$

where $P(\rho_k \pm \delta_k, t)$ indicates the configuration differing for the state of one node with degree k , the first two sums in the right hand side indicate the system is reaching the configuration $\boldsymbol{\rho}$, while the last one indicates the departure from the configuration. Making a Taylor expansion with respect δ_k of equation (5) till the second order, we obtain the Fokker-Planck equation for the system:

$$\begin{aligned} \delta t \frac{\partial P(\boldsymbol{\rho}, t)}{\partial t} &= \sum_k \frac{1}{N n_k} \frac{\partial}{\partial \rho_k} ((L(\rho_k) - R(\rho_k)) P(\rho_k, t)) \\ &+ \sum_k \frac{1}{2(N n_k)^2} \frac{\partial^2}{\partial \rho_k^2} ((L(\rho_k) + R(\rho_k)) P(\rho_k, t)) \end{aligned} \quad (6)$$

The coefficients in the sums of (6) can be expressed in terms of the quantities ρ_k, n_k and ω_2 as:

$$\begin{aligned} (R_k(\rho_k) - L_k(\rho_k)) &= n_k(\omega_2 - \rho_k) \\ (R_k(\rho_k) + L_k(\rho_k)) &= n_k(\rho_k + \omega_2 - 2\rho_k \omega_2) \end{aligned} \quad (7)$$

Moreover we notice that since $\delta_k^2/\delta t = 1/(N n_k^2)$, the second term in (6) is sub-leading and can be ignored, giving:

$$\frac{\partial P(\boldsymbol{\rho}, t)}{\partial t} = \sum_k (\omega_2 - \rho_k) P(\rho_k, t) \quad (8)$$

We use equation (8) to evaluate the time-evolution of the average value (on the ensemble of all the possible configurations $\boldsymbol{\rho}$) of ω_2 (indicated as $\langle \omega_2 \rangle$):

$$\begin{aligned} \langle \omega_2 \rangle &= \int \sum_k \frac{k^2 n_k \rho_k}{\mu_2} P(\boldsymbol{\rho}, t) d\boldsymbol{\rho} \\ \frac{d\langle \omega_2 \rangle}{dt} &= \int \sum_k \frac{k^2 n_k \rho_k}{\mu_2} \frac{dP(\boldsymbol{\rho}, t)}{dt} d\boldsymbol{\rho} \\ &= \sum_{k, k'} \int \frac{k^2 n_k \rho_k}{\mu_2} \frac{\partial((\rho'_k - \omega_2) P(\boldsymbol{\rho}, t))}{\partial \rho'_k} d\boldsymbol{\rho} \\ &= - \int \sum_{k, k'} \frac{k^2 n_k ((\rho_k - \omega_2) P(\boldsymbol{\rho}, t))}{\mu_2} \frac{\partial \rho_k}{\partial \rho'_k} d\boldsymbol{\rho} \\ &= \langle \omega_2 \rangle - \langle \omega_2 \rangle = 0 \end{aligned} \quad (9)$$

where we have integrated by parts and exploited the fact that the derivative term $\frac{\partial \rho_k}{\partial \rho'_k} = \delta(k, k')$, i.e., it is null when $k \neq k'$ and equal to 1 otherwise. The result implies that the average weighted magnetization is conserved for a fixed initial condition on the distribution $\boldsymbol{\rho}$. The existence of a conserved quantity, in our case ω_2 , determines the exit probability, the probability of reaching a consensus state [24]. Moreover we notice that the conservation of the average weighted magnetization determines the evolution of the density ρ_k :

$$\begin{aligned} \langle \rho_k \rangle &= \int \rho_k P(\boldsymbol{\rho}, t) d\boldsymbol{\rho} \\ \frac{d\langle \rho_k \rangle}{dt} &= \int \rho_k \frac{dP(\boldsymbol{\rho}, t)}{dt} d\boldsymbol{\rho} \\ &= \int \rho_k \frac{\partial((\rho_k - \omega_2) P(\boldsymbol{\rho}, t))}{\partial \rho_k} d\boldsymbol{\rho} \\ &= \langle \omega_2 \rangle - \langle \rho_k \rangle \end{aligned} \quad (10)$$

that has as solution:

$$\langle \rho_k(t) \rangle = \langle \omega_2 \rangle - (\langle \omega_2 \rangle - \rho_k(0)) e^{-t} \quad (11)$$

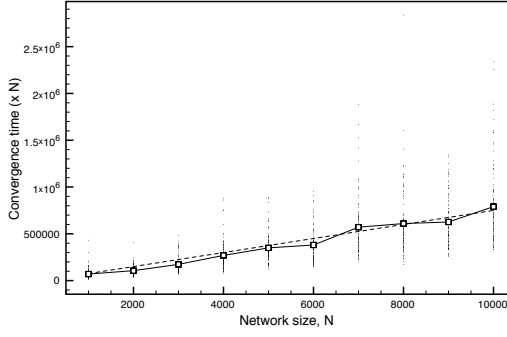


Figure 1: A simulation to show time to convergence for varying network sizes. We ran 100 independent trials for each network size, and the dots show the convergence time on each trial. The solid line shows the average convergence time. The dashed line plots $y = 75x$, which shows a very good fit with the numerically determined average convergence time.

meaning that immediately all the ρ_k reach the common value $\langle \omega_2 \rangle$. These last two results (9,11) (the weighted magnetization conservation and the time behaviour of $\langle \rho_k \rangle$) are valid for a generic value of β . The consensus time $T(\rho)$, the time at which all the nodes in the system have the same state, can be easily evaluated using the adjoint of the Fokker Planck equation (6):

$$\sum_k (\omega_2 - \rho_k) \frac{\partial T(\rho)}{\partial \rho_k} + \frac{1}{N} \sum_k (\omega_2 + \rho_k - 2\omega_2 \rho_k) \frac{\partial^2 T(\rho)}{\partial \rho_k^2} = -1 \quad (12)$$

Since $\rho_k \simeq \omega_2$ the first sum in (12) is null and can be eliminated. Moreover we can apply a change of variable:

$$\frac{\partial}{\partial \rho_k} = \frac{\partial \omega_2}{\partial \rho_k} \frac{\partial}{\partial \omega_2} = \frac{k^2 n_k}{\mu_2} \frac{\partial}{\partial \omega_2} \quad (13)$$

and then equation (12) can be rewritten as:

$$-1 = \frac{\partial^2 T}{\partial \omega_2^2} \left(\sum_k \frac{k^4 n_k}{N \mu_2^2} \omega_2 (\omega_2 - 1) \right) \quad (14)$$

The equation can be easily solved in terms of ω_2 :

$$T(\omega_2) = N \frac{\mu_2^2}{\mu_4} \left[(1 - \omega_2) \ln \left(\frac{1}{1 - \omega_2} \right) + \omega_2 \ln \left(\frac{1}{\omega_2} \right) \right] \quad (15)$$

The time of consensus depends on the initial randomness in the distribution of state through ω_2 , a finite term, and the size of network N explicitly and through the momenta. The size dependence is a function of both the exponent of the degree distribution and the momenta considered. The consensus time is a function of the momenta of the degree distribution and depends both on the exponent of the degree distribution ν and on the weight β . We consider $2 \leq \nu < 3$, the maximum degree being $k_{max} = N^{\frac{1}{\nu-1}}$, and for a generic m -momentum it follows that :

$$\mu_m \sim \begin{cases} N^{\frac{m}{\nu-1}-1} & m > \nu - 1, \\ \log N & m = \nu - 1, \\ 0(1) & m < \nu - 1. \end{cases} \quad (16)$$

The exponents of the momenta appearing in (15) are 2 and 4, which are larger than $\nu - 1$. Using (16), thus,

$$T(\omega_2) \simeq \text{constant} \quad (17)$$

meaning that the time to convergence is constant in the size of the network. Note that in this analysis, one time step is taken to involve N node updates. If we count each node update as a time step, then we expect a linear relationship between the size of the network and the time to convergence.

We verify the result numerically by generating random scale-free networks (without loners) with varying N and $\nu = 2.5$, and measuring the time to convergence. This is shown in fig. 1 where we plot the time to convergence for 100 independent trials for each network size. The network size was varied from 1000 to 10000 in steps of 1000. The convergence time for each run is plotted with a dot, and the average for each network size is shown with squares joined by a solid line. We see that a linear function, as expected, provides a good fit to the data.

4. MODELING INNOVATION

The DBVM assumes that the population begins with a set of variants, and no new variants are introduced after that. This raises the question, where do the original variants come from? One possibility is that for a given linguistic feature, only a few variants are possible, and they are found very quickly, leaving no room for further innovation of that feature. In this case, it is safe to say that all variants exist “from the beginning” in the population. This is not the case in certain instances of lexical change, where new words and near-synonyms for the same concept are not limited in numbers.

The other approach, then, is to say that some form of innovation is always occurring. In this case, there would be no reason to believe that only peripherals (or some other subgroup) innovate. We assume, instead, that anyone can innovate, at any time (though the innovation rate might be low). This view is close to the position adopted by Baxter et al., with their Utterance Selection Model [2], where nobody produces exactly the same utterance every time. In their model innovation can be understood as being due to random variation in speech production, or due to noise in the communication channel. In our model, however, we are interested in discrete innovation, i.e., we are modeling change in the lexicon, which may be triggered by external circumstances, such as the need for new words with the spread of new technologies, or increased contact between different speech communities.

It has been suggested, for instance, that in times of accelerated cultural change quite a few new items with new meaning as well as new items with near-synonymous meanings to existing words can enter the lexicon. Such lexical innovations can come from two sources. The first means of lexical enrichment that can lead to lexical inflation over time is borrowing, which can arise even in situations of relatively superficial cultural contact (see Thomason and Kauffman’s borrowing scale [26]). The second way is a specific type of language-internal innovation and borrowing process, called argots, jargons, and taboo [13, p. 420]. This second type seems to be the most appropriate analogy to consider with our innovation and diffusion model.

So the question we now ask is, during periods of high innovation rate, what sorts of norms will emerge in a population?

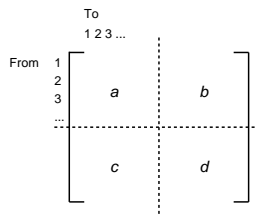


Figure 2: In the R-MAT algorithm, the adjacency matrix is recursively divided into quarters, and each quarter has a probability (a, b, c, d) associated with it. Starting with an empty matrix, we choose quarters recursively according to these probabilities until we get to a single cell, whereupon we set that cell to 1 to indicate a link.

We extend the DBVM to try to answer this question, by introducing a parameter p , the probability for innovation. In this model, a node can copy a neighbor chosen with probability proportional to the neighbors degree, as before, or, with probability p , introduce a new variant into the language.

We study the above question by keeping track of the source of each new variant, so that when we see a norm emerge in the population, we can tell which agent introduced it in the population. More precisely, we evaluate the probability that a variant that becomes a norm was introduced by a peripheral agent (or equivalently, by a non-peripheral agent).

5. THE DBVM WITH INNOVATION

In this extended model, we assume that there are v possible *initial* variants of a certain linguistic feature. To initialize the model, we assign a uniformly randomly chosen variant to each agent in the network at time $t = 0$. At each time step after that, we choose one of the agents uniformly randomly. This agent updates its variant by copying one of its neighbors with probability $(1 - p)$, where p is the innovation rate. With probability p , therefore, the agent introduces a new variant into the population. Variants are numbered starting with 1.

We keep a running count of the number of agents with each variant in the network. If one of the variants is in use by more than 90% of the population, we say that that variant has become the *norm*. Note that this means there can be periods when there is no norm in the population.

We also keep track of which agent introduced a particular variant, which will allow us to estimate the probability of variants generated by a particular class of nodes (e.g., loners) to become the norm in the network.

5.1 Generating the interaction network

Following Fagyal et al. [9], we generate the interaction network using the R-MAT algorithm [6]. R-MAT, which stands for Recursive MATrix, works by creating a set of nested communities in the network. The algorithm operates on the adjacency matrix of the network. An adjacency matrix describes a network as follows: if agent x is influenced by agent y in the social network (i.e. there is a link from x to y), then we place a 1 at row x and column y of the adjacency matrix, otherwise we place a 0 at that location.

The R-MAT algorithm uses four parameters, (a, b, c, d) , which correspond to four quarters of the adjacency matrix,

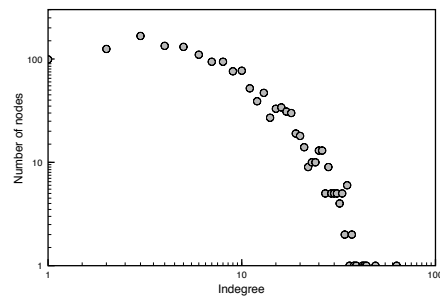


Figure 3: Indegree distribution of a network with 900 nodes and 7561 edges, generated by R-MAT.

as shown in fig. 2. We start with an adjacency matrix filled with zeroes. We then choose a quarter of the matrix with probability corresponding to its parameter. We chose the parameters $a = 0.5$, $b = 0.1$, $c = 0.1$, and $d = 0.3$. These parameters mean, for example, that half the time we choose the upper left quarter of the matrix. We then treat the chosen quarter as a new matrix, divide it into quarters, and again choose one quarter with the same set of probability parameters. This process is repeated recursively until we end up with a single cell, whereupon we set the value at that cell to 1. Again, following Fagyal et al. [9], we created a network with 900 nodes and added links to the adjacency matrix 9000 times, which resulted in 7561 unique links.

Another advantage of using the R-MAT algorithm is that it automatically results in a small number of loners ($\sim 5\%$ of the nodes), which avoids having to artificially choose a small number of peripheral nodes to designate as loners. The generated network has a heavy-tailed power-law-like degree distribution, as shown in fig. 3, and the behavior of the DBVM on these networks is similar to its behavior on scale-free networks.

6. SIMULATIONS

A single time step of the model corresponds to a single agent updating its variant, either by copying a neighbor or by innovating. Note that if an agent chooses to copy a neighbor, its variant may not actually change, because the chosen neighbor's variant might be the same as the agent's own.

Each simulation is run for 40 million time steps. We always start with $v = 8$ initial variants. The choice of number of initial variants is arbitrary; the qualitative dynamics are the same for other (small) values of v . Once the simulation starts, agents introduce new variants in the population with innovation rate p . We count the number of individuals for each variant in the population every ten thousand time steps. If a particular variant is being used by more than 90% of the population, we say that it is the norm. We mark this on the graph by a single point for that variant number at that timestep.

Figure 4 shows norms when the innovation rate, $p = 0.0001$. We see that nearly all the time, the norm is one of the original eight variants (which are numbered from 1 to 8). Very rarely, a new variant (with number greater than 8) becomes the norm.

Figure 4 suggests that if we observe a variant as the norm in a population, it is due to a peripheral member, with high probability. The next simulation increases the innovation

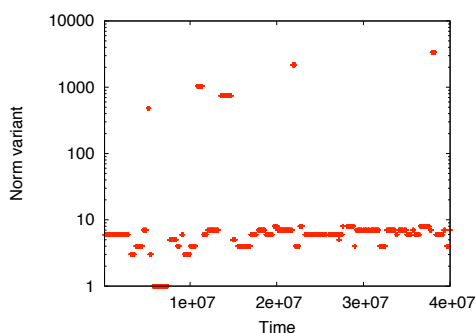


Figure 4: Norms are primarily the variants held by peripheral members, when the innovation rate is 0.0001.

rate by an order of magnitude, i.e., $p = 0.001$ now. The result is shown in figure 5. We see that, in this case, norms are much more evenly split between the original variants and innovative variants. This shows that probability of a non-peripheral-introduced variant becoming the norm depends on the innovation rate. In other words, if we observe a norm in a network, the statistical answer to “who introduced this into the population?” depends on the rate at which innovations are being introduced into the population as a whole. To get a more precise picture, we did a number of runs for various values of p , varying it from 0.0001 to 0.01. The results are shown in figure 6.

We did a ten runs for each value of p . Figure 6 shows two curves. The dashed line is the average fraction of the total simulation time for which a norm exists in the population. We call this the *norm time*. The norm time varies from one run to another because, even though the network is the same every time¹, the initial state of all the nodes is set randomly. The solid line shows the fraction of the norm time for which the norm was a variant introduced into the population by a loner. We call this the *loner fraction*. The error bars show one standard deviation.

Note that while the total number of variants generated over the span of the simulation is quite large, there are relatively few variants circulating in the network at any given time. The lifetime of an innovation is quite short because new variants are lost with high probability as nodes re-copy an existing variant from another node after they generate an innovation.

There are a few interesting things to note in figure 6. One is that as the innovation rate increases, the fraction of time that the norm is a peripheral-introduced variant decreases and correspondingly the fraction of time that the norm is a non-peripheral-introduced variant ($1 - \text{loner fraction}$, not shown in figure 6) increases. Second, at the same time, the fraction of the total time for which a norm exists at all in the population decreases with increasing innovation rate. When the innovation rate is 0.01, i.e. when an agent innovates only with a one in *hundred* probability, no norms appear in the population. This means that agents have to be *rather* conservative if norms are to exist at all. Third, we can use the fraction of time that a norm exists at all in the popu-

¹Since we use only one network, the values we have computed are network-specific, but the qualitative results are the same across different network instantiations.

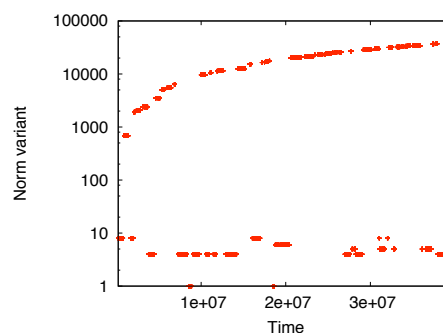


Figure 5: Norms are relatively equally divided between peripheral and non-peripheral variants when the innovation rate is 0.001.

lation as an index to determine innovation rate, and thus the probability that the norm is due to a peripheral member. This means that even if we do not know the rate at which innovations are being introduced into the population (and empirically, we can’t), we can still estimate the probability that a norm is due to an innovation introduced by a peripheral.

As the innovation rate increases, the loner fraction decreases, which means that it becomes more and more likely that an innovation introduced by someone other than a loner can become the norm. The loner fraction drops below 0.5 when the innovation rate is approximately 0.002 in this simulation. At this point, it becomes more likely that an innovation introduced by a non-loner will become the norm, than that an innovation introduced by a loner will become the norm. Note that for this value of the innovation rate (and above this value), the norm time has dropped to about 25% or less. Thus, for variants introduced by non-loners to be more likely to become the norm, the innovation rate must be so high that norms only exist in the population for brief intervals.

7. LEXICAL INFLATION IN FRENCH

These findings seem to align with certain types of lexicosemantic change, such as lexical inflation, in natural languages. The following examples will focus on lexical change in French, which corresponds to one of the best known and described examples of this type of change in modern European languages. Lexical inflation is a process by which lexical items with the same meaning and similar stylistic use tend to accumulate and persist in the lexicon over time [21, p. 155], [10, p. 118]. There is general consensus among linguists that the lexicon resists the inclusion of too many perfect synonyms, i.e. lexical items duplicating the same meaning, but partial or near-synonyms can be quite numerous. While theoretical models of near-synonymy are still debated (see [25]), their practical implications have been observed for many years.

Parallel to new near-synonyms entering the language, old lexical items also need to persist for lexical inflation to occur. As Posner [21, p. 155] notes with respect to lexical inflation in French: “Most words that have outlived their time are not consigned to the dustbin, but to the attic, whence they can be taken out, dusted down, and brought back into use for special occasion.” In other words, words do not neces-

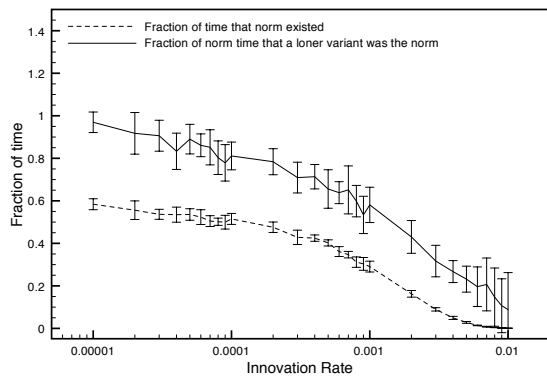


Figure 6: The dashed line shows the fraction of the total time for which a norm existed in the population, which we call “norm time”, against innovation rate. We see that the norm time becomes essentially zero when the innovation rate is close to 0.01. The solid line shows the fraction of *norm* time for which the norm was a variant introduced by a loner (or one of the original variants, which are held on to only by loners after the first norm emerges). We call this the “loner fraction”.

sarily disappear; just become more peripheral in the lexicon. Word-loss has indisputably characterized French, and other languages, historically. An especially large number of words were lost from French during the early modern period. Verb forms, such as *gésir* (to lie) and *quéirir* (to seek) that, although not particularly frequent, had particularly irregular conjugations led to the replacement of these verbs by regular forms (*coucher* (to lie), *chercher* (to seek)). However, it has also been argued that since at least the late 18th-century the overall picture seems to be that of lexical inflation: “the number of different words (types) occurring in texts has not ceased to grow, even though for some individual words the number of instances (tokens) has regressed.” [21, p. 155]. The question is: what possible mechanisms, if any, might have motivated this type of inverse relationship between type-frequency and token-frequency?

Furthermore, as near-synonyms tended to accumulate and persist in the lexicon starting from the late 18th-century, historians of French also noticed that the use of many previously peripheral and/or specialized lexical items became generalized. Casting his net of spoken lexical forms much wider than dictionaries of his time, Lazare Sainéan [22], among others, left literary usage behind to analyze an impressive array of fringe vocabulary spoken by the early 20th-century Parisian society’s have-nots. He studied the jargon of soldiers, butchers, sailors, shoemakers, printers, and other corporations, as well as the secret terms, or argots, of thieves, beggars, prostitutes, pimps, and professional gamblers. Together with the terms of his times’ entertainment industry, the Parisian cabarets, Sainéan also listed the meaning and stylistic connotations of terms handed down from child language and dialectal borrowings, or provincialisms. This wide variety of lexical items studied in their historical context lead him to one general conclusion: “Having followed the evolution of the language of criminals until the 19th-century, I came to the conclusion that the last traces of

this idiom (whose sole reason to exist was its secrecy) have blended into modern-day working-class Parisian French. [...] This modern argot led to a unified idiom spoken by millions of Parisians and French people.” [22, VII-VIII]. The sole reason for this “penetration of jargon into ‘the vulgar’ (i.e., working-class spoken French)”, according to Sainéan, was the result of more frequent and “infinitely more easy” contact between different segments of French society.

Analogies between the dynamics of our computational model and the above story of lexical innovations in industrial-age Paris are suggestive. Increased innovation and large-scale spread of slang words and group-specific technical terms are first noticed in French in the modern era, i.e. starting from the late 18th-century when task-oriented labor divisions and technological advances in the manufacturing sector bring in close and regular contact members of traditionally tight-knit communities in close-reach from each other (i.e., small in diameter and showing high clustering). Marked by flagrant social inequalities (i.e., possibly of scale-free degree distribution), these networks could have been prominent sites of the type of innovation and distribution dynamics exhibited in our general model. The question whether the (inverse) relationship between increased type-frequency and decreased token-frequency is indeed governed by the same statistical dynamics as the innovation rate increase vs. norm-time decrease in our model remains to be investigated empirically in very large written and spoken language corpora. What we hope to have accomplished in this paper is a more precise formulation of the next series of hypotheses to be tested on lexical inflation in French and other languages.

8. CONCLUSION

In this paper, we have analyzed and extended a model of linguistic innovation and diffusion in social networks. We have shown how to derive the time to convergence in the degree-biased voter model. Our analysis follows the technique of Sood and Redner [24] of grouping nodes by degree to derive the Fokker-Planck equation for the system. From this we derive the adjoint equation, and the expression for convergence time follows. It turns out that time to convergence in the DBVM is simply linear in the size of the network, when time is measured as the number of updates, which we confirmed with a simulation.

The previous model is analogous to stable sedentary societies where there are a small number of variants for any linguistic variable. However, as is well-attested in historical linguistics, during periods of accelerated cultural change, languages must adapt to a greater number of innovations, especially in the lexicon. We model this situation by including a probability of innovation into the DBVM. We did simulations to qualitatively understand the nature of this extended model, and saw that as innovation rate increases, the duration of norms decreases, as is indeed the case historically. We also discovered that the probability of loner or peripheral variants becoming the norm tends to be substantially higher than non-loner variants. This has also been empirically noted, in 19th-century French for example, which saw a large number of terms from argots and jargons being incorporated into the mainstream. Our approach suggests that a simple stochastic model might account for a great deal of this change.

We do not, however, claim that the above are the only reasons for linguistic change, or that simple stochastic models

can account for all the variation observed empirically. There are a number of essential sociolinguistic factors left out by our model, including effects of gender, age, and social identity. Our goal is to model these factors incrementally, in order to make sure that the effects of each new factor are fully examined before including them in the model.

We end this paper by underscoring the importance of computational modeling in sociolinguistics. Language is a very complex adaptive system. The dynamics of large-scale interactions and long-terms change are, we believe, impossible to understand fully without a rigorous mathematical theory and computational tools [17] that allow linguists to experiment with factors identified in small-scale empirical studies in sociolinguistics.

9. ACKNOWLEDGMENTS

S.S. is supported in part by NSF Nets Grant CNS-0626964, NSF HSD Grant SES-0729441, NSF PetaApps Grant OCI-0904844, NIH MIDAS project 2U01GM070694-7, DTRA R&D Grant HDTRA1-0901-0017, DTRA CNIMS Grant HDTRA1-07-C-0113, NSF NETS CNS-0831633, DHS 4112-31805, DOE DE-SC0003957, NSF REU Suppl. CNS-0845700, US Naval Surface Warfare Center N00178-09-D-3017 DEL ORDER 13, NSF Netse CNS-1011769 and NSF SDCI OCI-1032677.

A.A. is supported by DynaNets. DynaNets acknowledges the financial support of the Future and Emerging Technologies (FET) program within the Seventh Framework Program for Research of the European Commission, under FET-Open grant number: 233847.

Zs.F.'s research on this topic is supported by the University of Illinois's Research Board for the project entitled "Modeling lexical change the role of political centers and marginal groups in the selection and spread of lexical innovations in 19th- and 20th-century France."

10. REFERENCES

- [1] G. Baxter, R. Blythe, W. Croft, and A. McKane. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, 21(2):257–293, 2009.
- [2] G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane. Utterance selection model of language change. *Physical Review E*, 73:046118, 2006.
- [3] C. Beckner, R. Blythe, J. Bybee, M. H. Christiansen, W. Croft, N. C. Ellis, J. Holland, J. Ke, D. Larsen-Freeman, and T. Schoenemann. Language is a complex adaptive system: Position paper. *Language Learning*, 59:1–26, 2009.
- [4] L. Bloomfield. *Language*. U. Chicago Press, Chicago, London, 1933.
- [5] D. Britain. Exploring the importance of the outlier in sociolinguistic dialectology. In D. Britain and J. Cheshire, editors, *Social Dialectology: In Honour of Peter Trudgill*, pages 191–208. Amsterdam/Philadelphia: John Benjamins, 2003.
- [6] D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-MAT: A recursive model for graph mining. In *SIAM conference on data mining*, 2004.
- [7] P. Eckert. Adolescent social structure and the spread of linguistic change. *Lang. Soc.*, 17(2):183–207, June 1988.
- [8] P. Eckert. *Linguistic Variation as Social Practice*. Blackwell, Malden, MA, 2000.
- [9] Z. Fagyal, S. Swarup, A. M. Escobar, L. Gasser, and K. Lakkaraju. Centers and peripheries: Network roles in language change. *Lingua*, 120(8):2061–2079, 2010.
- [10] R. Gouws. Aspects of lexical semantics. In *Solving Language Problems: From General to Applied Linguistics*, pages 98–132. Exeter University Press, Exeter, 1996.
- [11] M. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.
- [12] M. Granovetter. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1:201–233, 1983.
- [13] H. Hock. *Principles of Historical Linguistics*. Mouton de Gruyter, Berlin, New York, 2nd edition, 1999.
- [14] E. Katz and P. Lazarsfeld. *Personal Influence*. Free Press, Glencoe, IL, 1955.
- [15] W. Labov. The social origins of sound change. In W. Labov, editor, *Locating Language in Time and Space*, pages 251–266. Academic Press, NY, 1980.
- [16] D. Lønsmann. From subculture to mainstream: The spread of English in Denmark. *Journal of Pragmatics*, 41:1139–1151, 2009.
- [17] 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(6014):176–182, January 2011.
- [18] J. Milroy and L. Milroy. Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21:339–384, 1985.
- [19] L. Milroy. *Language Change and Social Networks*. Blackwell Publishers, Oxford, 1987.
- [20] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.
- [21] R. Posner. Lexical change. In *Linguistic Change in French*, pages 143–184. Clarendon Press, 1997.
- [22] L. Sainéan. *Le Langage Parisien au XIXe siècle: facteurs sociaux, contingents linguistiques, faits sémantiques, et influences littéraires [Parisian French in the 19th Century: social factors, linguistic elements, semantic facts, and literary origins]*. Champion, Paris, 1912. Available online: <http://gallica.bnf.fr/ark:/12148/bpt6k5433966k>.
- [23] V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Phys Rev E*, 77:041121, 2008.
- [24] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94:178701, 2005.
- [25] J. R. Taylor. Near-synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences*, 25(3):263–284, 2003.
- [26] S. G. Thomason and T. Kaufman. *Language Contact, Creolization, and Genetic Linguistics*. University of California Press, Berkeley, 1988.
- [27] U. Weinreich, W. Labov, and M. I. Herzog. Empirical foundations for a theory of language change. In W. P. Lehmann and Y. Malkiel, editors, *Directions for Historical Linguistics: A Symposium*, pages 95–188. University of Texas Press, Austin, 1968.